

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

Protein - protein search, using sw model

Search on: January 22, 2001, 11:50:43 ; Search time 233.01 Seconds
(without alignments)
223.791 Million cell updates/sec

Title: US-09-540-245A-2
E-value score: 8316
Sequence: 1 MRGCVQWMLSLSLGLVLAIL.....SSFYDEVKVKYCGKTCRVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues ;
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8316	100.0	1525	20 Y17499	Human Slit-1 prote
2	8265	99.4	1529	20 Y27145	Human slit-2 prote
3	8265	99.4	1529	20 W96702	Full length slit-1
4	8137	97.8	1503	20 T27142	Human slit-2 matur
5	8137	97.8	1503	20 W96701	Slit-like protein
6	8065	97.0	1529	21 Y76117	Rat Slit homologue
7	5717.5	68.8	1523	21 Y99395	Human PRO1336 (UNQ
8	5714.5	68.7	1523	20 Y27146	Human slit-3 prote
9	5714.5	68.7	1523	20 Y04137	Human slit 3 prote
10	5702.5	68.6	1496	20 Y27143	Human slit-3 matur
11	5702.5	68.6	1496	20 Y04136	Human slit 3 matur
12	5611.5	67.5	1523	20 Y14142	Human Slit protein

13	5597	67.3	1534	19 W46966	Amino acid sequenc
14	5589	67.2	1534	20 Y27144	Human slit-1 prote
15	5589	67.2	1534	20 Y04139	Human slit 1 prote
16	5589	67.2	1534	20 W96707	Protein sequence o
17	5588	67.2	1508	20 Y27141	Human slit-1 matur
18	5588	67.2	1508	20 Y04138	Human slit 1 matur
19	5588	67.2	1508	20 W96706	Protein sequence o
20	3543.5	42.6	716	21 Y76055	Rat slit homologue
21	3443	41.4	1480	13 R25079	Drosophila Slit pr
22	1658	20.1	299	20 Y27149	Protien encoded by
23	1658	20.1	299	20 W96703	EST clone protein
24	1429.5	17.2	450	20 Y27156	Peptide Seq ID No:
25	766	9.2	1010	20 W87896	Human JAGGED1 solu
26	766	9.2	1187	18 W18352	Proliferation and
27	766	9.2	1218	17 W05833	Human Serrate-1 (H
28	766	9.2	1218	19 W44301	Human serrate 1.
29	766	9.2	1218	20 W87894	Human JAGGED1 prot
30	766	9.2	1218	21 Y59597	Human Serrate prot
31	761	9.2	1036	18 W18351	Proliferation and
32	761	9.2	1218	18 W18354	Proliferation and
33	759	9.1	1208	19 W40827	Human Jagged prote
34	755	9.1	228	19 W46967	Amino acid sequenc
35	755	9.1	228	20 Y27147	Mouse slit protein
36	737.5	8.9	2471	20 Y06816	Human Notch2 (humN
37	736.5	8.9	1193	17 W05835	Chick Serrate. Ga
38	736.5	8.9	1193	21 Y59599	Human Notch3 prote
39	719.5	8.7	2321	19 W49698	Human Notch3 prote
40	697	8.4	1872	19 W68510	Partial human Notc
41	691.5	8.3	1964	20 W95557	Mus musculus notch
42	685	8.2	1148	20 W87895	Human JAGGED2 prot
43	680.5	8.2	1055	19 W44298	Human serrate 2 pr
44	680.5	8.2	1212	19 W44299	Human serrate 2
45	680.5	8.2	1257	17 W05834	Human Serrate-2 (H

ALIGNMENTS

RESULT 1

Y17499

ID Y17499 standard; Protein; 1525 AA.

XX

AC Y17499;

XX

DI 04-AUG-1999 (first entry)

XX

DE Human Slit-1 protein.

XX

KW Human; Slit-1; Robo; modulation; identification; interaction.

XX

OS Homo sapiens.

XX

PX WC9925811-A2.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24245.

XX

PR 07-APR-1998; 9805-0081057.

XX

PR 14-NOV-1997; 9705-0065544.

XX

PA (RSC) UNIV CALIFORNIA.

XX

PI Brose K, Goodman C, Kid T, Tessier-Lavigne M;

XX

DR WPI; 1999-347475/29.

XX

DR N-PSDB; Y76161.

XX

PT Human Slit polypeptide and related nucleic acids

XX

PS Disclosure; Page 19-21; 34pp; English.

XX

CC The present sequence is the human Slit-1 protein. The present invention

CC also describes a method for identifying agents which modulate the
 CC interaction of Robo and a Robo ligand comprising: combining a Robo
 CC polypeptide, a Slit polypeptide and a candidate agent under conditions
 CC where the Robo and Slit polypeptides normally (but for the presence of
 CC the agent) engage in a first interaction, where the Slit polypeptide
 CC specifically binds, activated or inhibits the activation of the Robo
 CC polypeptide and determining a second interaction of the Robo and Slit
 CC polypeptides in the presence of the agent, where a difference between
 CC the first and second interactions indicates that the agent modulates the
 CC interaction of the Robo and Slit polypeptides; and a method to modulate
 CC the interaction of Robo and a Robo ligand. The method is useful for
 CC screening for Robo (roundabout) modulators and Robo:Slit complexes are
 CC useful for regulating various cell functions, especially of neuronal
 CC cells.

XX
 50 Sequence 1525 AA;

Query Match 100.0%; Score 8316; DB 20; Length 1525;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRGSGWQMSLSGLWLAALNVAQACPAQCSGSGSTVDCGSLALSVFNPANPERXK 60
 Db 1 MRGSGWQMSLSGLWLAALNVAQACPAQCSGSGSTVDCGSLALSVFNPANPERXK 60
 Qy 61 DLNMGNTNTRITKTFAGLRHLRLVQLMENKSTIERGAFODLKEHLRLNMRHLQFPE 120
 Db 61 DLNMGNTNTRITKTFAGLRHLRLVQLMENKSTIERGAFODLKEHLRLNMRHLQFPE 120
 Qy 121 LFLFGTAKLRLLDLSNQIQAIKRAFGAVDINKLQDLSQISCEGAFRALRDEVL 180
 Db 121 LFLFGTAKLRLLDLSNQIQAIKRAFGAVDINKLQDLSQISCEGAFRALRDEVL 180
 Qy 181 TLNNNTNTRLSVAFSNMMPKRTFFRLHSNNLTCDOHLASDMLKRPVGLTQCMQPS 240
 Db 181 TLNNNTNTRLSVAFSNMMPKRTFFRLHSNNLTCDOHLASDMLKRPVGLTQCMQPS 240
 Qy 241 HLAGHNVAVQKREFVCSDEEGHGFMAFPCSVLHCACTCSNNIVLCRGSLTEIPT 300
 Db 241 HLAGHNVAVQKREFVCSDEEGHGFMAFPCSVLHCACTCSNNIVLCRGSLTEIPT 300
 Qy 301 NLPEITAIRLBNQITKIVPPGAFSPYKLRRLDLSNQISLAPADQGLRSLNSLVLY 360
 Db 301 NLPEITAIRLBNQITKIVPPGAFSPYKLRRLDLSNQISLAPADQGLRSLNSLVLY 360
 Qy 361 GNKITELPKSLFEGFLSQLLLLNANKINCLRVDAFDLHNLNLSLYDNKLTQAKGT 420
 Db 361 GNKITELPKSLFEGFLSQLLLLNANKINCLRVDAFDLHNLNLSLYDNKLTQAKGT 420
 Qy 421 SPLRAIQTMHLAQNPFIDCHLKWADYLHNPITSGARTCSPRRLANKRIGQLSKKF 480
 Db 421 SPLRAIQTMHLAQNPFIDCHLKWADYLHNPITSGARTCSPRRLANKRIGQLSKKF 480
 Qy 481 RCGSTEDYRSKLSGDCFADLACEKRCRCSTVTVCSNQKLNKPIPHQPTAELRLNNE 540
 Db 481 RCGSTEDYRSKLSGDCFADLACEKRCRCSTVTVCSNQKLNKPIPHQPTAELRLNNE 540
 Qy 541 PTVLEATGFIKKLQPLRKINFSNNKITDIEGAFEGASGVNELLTSNRLNVQHMFKG 600
 Db 541 PTVLEATGFIKKLQPLRKINFSNNKITDIEGAFEGASGVNELLTSNRLNVQHMFKG 600
 Qy 601 LESLAKTLMSNRITCVGNSPIGLSSVRLLSLYDNQITVAGAFDTLSLSTLALLAN 660
 Db 601 LESLAKTLMSNRITCVGNSPIGLSSVRLLSLYDNQITVAGAFDTLSLSTLALLAN 660
 Qy 661 PFNCCNYLAWLGELWKRKRIVTGNPRQCFYFLKEIPIDVQIDQITCDGNDNSCSPL 720
 Db 661 PFNCCNYLAWLGELWKRKRIVTGNPRQCFYFLKEIPIDVQIDQITCDGNDNSCSPL 720
 Qy 721 SRCPTCTCLDTVVRCSNKGVLKPKIPDVTELYDGNQFTLVKPSLNTYKHLIDL 780
 Db 721 SRCPTCTCLDTVVRCSNKGVLKPKIPDVTELYDGNQFTLVKPSLNTYKHLIDL 780

Qy 781 SNKRISTLSNQFSNNQTLTLILSYNRLRCIPPRFDGLKSLRLLSLHGNDSIVVPEGA 840
 Db 781 snkristslseqsfnnqtltililsynrlrcipprfdglsrlslhgnsvvpega 840
 Qy 841 FNDLSLHLSHLAGNAPLYCDNQWQLSDNVKSEYEPGIRACAGPGMAOKLLITTPSKK 900
 Db 841 fndlsllshlagnaplycdcnqwlsvdwkeyeppgllaracagpgmadkllittpskk 900
 Qy 901 FTCQGGPDWNLAKNCPNLSNCKMGKCSNDPVDYRCTCPYCPKQGDVPIHACISN 960
 Db 901 ftcqggpdwnlakncplnsckmgkcsndpvdyrctcpcyckqgdvpihacisn 960
 Qy 961 PCKHSGTCHLAGEEDGFWICADGPEENCEVNWDDCEDNENNSTCVGINVYTCIC 1020
 Db 961 pckhsgtchlageedgfwicadgpeencevwnddcednennstcvginvnytcic 1020
 Qy 1021 PPEYTGELCEIKLFCQADLWPCQDSKCLILPKGCKTDPGVVGEHCIDIDFDQCNK 1080
 Db 1021 ppeytgelceekldfcadlnpcqdsckclitpkgtcdtpgvvgehcididfdqcnk 1080
 Qy 1081 CKNAGKATDANVGYTICPEGISGLCFEFSPPWLVPTSPCNDFCQGAQCVINAPNEI 1140
 Db 1081 cknagkatdavnvyticpegisglcfefspwlvptspcndfcqgaqcvinapnei 1140
 Qy 1141 CQCLPYGQKCEKRLVSNVFNKESYLIPASIKVRPQTNITLQIATEDESGILLIKGMD 1200
 Db 1141 cqclpygqkeckrlvsnvfnkesylipaskvrpqtntitlqiatdesgillikgmd 1200
 Qy 1201 HIAVELHGRVSRVASTDYGSHPSAIIYSVETINDGNHIVELLADQSLISVGGNPKII 1260
 Db 1201 hiavelhgrvsrvastdygshpsaaiysvetindgnhivelladqslisvggnpkii 1260
 Qy 1261 TNLKSQNTLNTSDPLTVGMPGKSNVASLQAQAGQNGTSFHCIRNLINSELQDQKVP 1320
 Db 1261 tnlksqntlnstsdpltvvgmpgksnvaslqaqagngtsfhcirliniselqdqkvp 1320
 Qy 1321 MDTGLPGCECHKVCNAGTQCPSSQAGTCECQGWGMLPCDQRTNDPLGNKRCVBT 1380
 Db 1321 mdtglpgcechkvcnagtcqpsqagtccecgwgmplcdqrdndplgnkrcvbt 1380
 Qy 1381 CLPINFASYSCKLEGRGVLCDEEDLFNPQAIKHKGRKLSGLQPLBCSSGTYG 1440
 Db 1381 clpinafsyscklegrgvlcdeedlfnpqaikchkgrklsqplbcssgtyg 1440
 Qy 1441 DSCDRISCRGERIDITYQKQGYAACQTTKVRSLRCRGGCAGGCGCPLASKRRKYSF 1500
 Db 1441 dscdriscrgeridityqkqgyaacqttkvrslrcrggcagcgccplaskrrkysf 1500
 Qy 1501 BCTDGSSFDVEKVRVKCQTRCVS 1525
 Db 1501 bctdgssfdvekvrvkcqctrvs 1525

RESULT 2

Y27145

ID Y27145 standard; protein; 1529 AA.

XX

AC Y27145;

XX

XX

DT 15-SEP-1999 (first entry)

XX

DE Human slit-2 protein (Seq ID No: 12 of JP11164690).

XX

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;

KW muscle; endocrine system.

XX

OS Homo sapiens.

XX

PN JP11164690-A.

XX

PD 22-JUN-1999.

XX

PF 05-DEC-1997; 97JP-0335435.

XX 05-DEC-1997; 97JP-0335435.

XX (ASAH) ASAH KASEI KOGYO KK.

XX WPI; 1999-411830/35.

DR N-PSDB: X89162.

XX
PT New vertebrate slit protein - useful for diagnosis and treatment of
PT cancers in nerves, muscle and endocrine system
XX
PS Disclosure; Page 58-63; 102pp; Japanese.

XX
CC The invention relates to a vertebrate-derived protein containing an amino
CC acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein
CC has at least 55 % homology to one of sequences shown in Y27141-Y27143,
CC and has slit protein-like activity. The vertebrate slit proteins encoding
CC nucleic acid sequences have at least 50% homology to nucleic acid
CC sequences X89161-X89168. The vertebrate-derived proteins can be produced
CC recombinantly by transfecting host cells with expression vectors
CC comprising the encoding nucleic acids. The proteins of the invention are
CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine
CC system.

XX Sequence 1529 AA;

Query Match 99.44; Score 8265; DB 20; Length 1529;
Best Local Similarity 99.24; Pred. No. 0;
Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 1 MRGVMQMLSLGLVLAAILNKVAPQACQACSCSGSTVDCBGLALSRVNPINRTERL 60

Db 1 mrqvgmqlslglvllaailnkvpagacqacscsgstvdcbglalsrvnprnterl 60

Qy 61 DLNMGNTIRITRTDFAGLRHLRVQLMENKSTIERGAFQDLKELRLNRLNHLFPE 120

Db 61 dlnmgntirritrtdfaglrhlrvqlmenkstiergafgdlkelerlrlnrlnhlfp 120

Qy 121 LLFGLTAKLYRDLSENQIQAIPKAFKAVDINKLQDYNQISCEIDGAFRALRDLVL 180

Db 121 llfgltaklyridlsenqiaipkafkavdinklqdyngiscedgafrairdl 180

Qy 181 TLNNNITRLSVASFENMPKLRITFLHSNNLYKCDLAWLSNWLKRPFWGLTQCMGSP 240

Db 181 tlannnitrslvasfennpkrlritflhsnnlykcdlawlsnwlkrpfrwgltycmgsp 240

Qy 241 HLHGHVAVQVQKREYFVCSDEEGHOSFMAPSCSVLPCAPACCSNNIVCGRGLTEIPT 300

Db 241 hlhghvavvqvkrefvcs---ghqgsfnpscvlhpcactcsnnivcgrglteipt 296

Qy 301 NLPETITRILRQNTIKVIPPQAFSPYKLRIDLSNQLISLAPQALRSLNSLVLY 360

Db 297 nlpetitrlrqnrtikvipppafspykrlridlsnqslslapqalrslnslvly 356

Qy 361 GNWITELPSLFEGLFSLQLLLNANKINCLVADQADLHNNLKLSDYDNKLTQIANGFT 420

Db 357 gnwitelpslfeqlfslqlllnankincivadqadlhnlnllydnkltqiangft 416

Qy 421 SPLRATQTHLQNPFFIDCHLKLWADYIHTNPITSGARCTSPRLNKRIGIKSKKF 480

Db 417 splratqthlqnpffidchklwadyihtnpitsgarctsprrlnkrigikskkf 476

Qy 481 RCS-----GTEDYRSKLSGDCFDALACEPKRCRGTVDCSNQKLNKIPHEIPQYTA 532

Db 477 rcsaegfyigtedyrsklsqdcfdalacepkrcrgtvydcsnqklnkipehpyta 536

Qy 533 ELRLNNNEPTVLEATGIFKKLPQLRKINFSNNKTDIEBAGFEGASGVNELLSTNLEN 592

Db 537 elrlnnnetvleatgiffkplqlrklnfsnnktdiebagfegagsgvnellstnlen 596

Qy 593 WQHKMGKLESKTLMLRSNRITCVGNDSFGISSLVRLSLYDNCITTVAGAFDTLHSL 652

|||||
Db 597 wqhmkgleskltmlrsnrictvgnndsfglssvrlsllydncitvtagafdtlhl 656

Qy 653 STLANLNPFWNCNYLANLGEWLKRRIVTNGRCQPKFKEIPIDQVADQPTDCDGN 712

Db 657 stlanlnpfnccnylawewlkrkrivtngprcqkyfkeiplqdvadqfctddgn 716

Qy 713 DDMSCPLSRCPTECTCLDVTVRCSNKLKVLPGIPRDVTELYLQGNPTLPKLSNY 772

Db 717 ddmcsplsrctpectcldvtrcsnklkvlpgiprdvteilydgnptlpklsny 776

Qy 773 KHLTLIDLSNNRISTLSNQSFSNMQLTLILSYNRLICPPRTFDGLSKLLSLHND 832

Db 777 khltilidlsnnrilstsqsfsmqltililsynrlrcpprtfdglskrlslhnd 836

Qy 833 ISVPEGAFNDLSALSHLAGANPLYCNCNQLSDWKSSEYKEPGIARCAGSGEMADKL 892

Db 837 isvpegaafndlsalshlaiganplycncnqlsdwksseykepgiarcagsgemadkl 855

Qy 893 LLTPSKKFTQCGPVNLIANKNCLNPKNKNDGTCNSDVFYRCTCPYFGKQGV 952

Db 897 lltpskkftqcgpvvnliankncplsnkndgtcnsdvpfyrctcpgyfgkqgv 956

Qy 953 PIHACISNPKRGGTCHLKEGEDFWICADGFEGENCEVNDCCEDENNSTCVDG 1012

Db 957 pihacisnpgkgtchlkgeedfwicadgfegecnvndcceddennstcvdg 1016

Qy 1013 INNYTCLCPPTETGELCEKLDQACDLPNCQDSCILTPKFGKDCDTPGVYGHCD 1072

Db 1017 innytclcpptetgclceklcdqadlpncqdsckiltpkgkdcctpgvyghcd 1076

Qy 1073 FDDCQDNKCNKAGRCTDAVNGYTCIPEGYSLGCEFSPPMVLPTSPCFNDCQAGAC 1132

Db 1077 fddcqdnkcnagactdavnngytcipegyslgcefsppmvlptspcfndcagagac 1136

Qy 1133 IVRINEPQICQLPGYQGECKELSVNFINKESYLQIPSAKVRPQTNITLQIATEDSGI 1192

Db 1137 ivrinepqcqlpgyqgeckelsvnfinkesylqipsakvrpqtntilqiatdsgl 1196

Qy 1193 LLYGKDHIAVELYRGRVRSYDTGSHPAASIVSYETINDGNFHIVELALDQSLSLV 1252

Db 1197 llygkdhiavelyrgrvrsydtgshpaasivsyetindgnfhivelaldqslslv 1256

Qy 1253 DGNPKIITNLKSQSTLNFDSPLYVGGMPGKSNVASLRQAPQNGTSGFCGRNLYNSE 1312

Db 1257 dgpnpkiitnlksqstlnfdsplyvggmpgksnvaslrqapngtsgfncgrnlyns 1316

Qy 1313 LQDFQKVPMTGILPGCEPCHKVKCAHGTCPSSQAGFTCEQCGNMPLCDQRTNDPL 1372

Db 1317 lqdfqkvpmtgilpgcepchkvkahgtcpssqagftceeqcgnmplcdqrdndpl 1376

Qy 1373 GNKVGKGLCLPAFYSYCKLBSGGVLCDEEDLFWQCAIKCKHGRKRLSGLGQVTC 1432

Db 1377 gnkvgtgclplafysycklbgggvlcdeedlfwqcaikckhgrkrlsglgqvtc 1436

Qy 1433 ECSSGFTGDSIDREISCRGERIDRYTQKQSTACQATTKVSHLBCRGGAGCGGLR 1492

Db 1437 ecsgsftgdsidreiscrgeridrytqkqstacqattkvshlbcrggagcgglr 1496

Qy 1493 SKRRKTSFECTDGSFDEVEKVVRCGCTRCVS 1525

Db 1497 skrrktsfctdgsfdevekvkvgctrcvs 1529

RESULT 3

W96702

ID W96702 standard; protein; 1529 AA.

XX

AC W96702;

XX

DT 15-APR-1999 (first entry)

XX

DE Full length slit-like protein sequence.

XX
 KW Slit-like polypeptide; diagnosis; treatment; nervous disease;
 KW thyroid disease; adrenal disease; muscular disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..26
 FT /note= "signal peptide"
 FT Protein 27..1529
 FT /note= "mature protein, claimed in Claim 1"
 XX
 PN JP11018777.1A.
 XX
 PD 26-JAN-1999.
 XX
 PF 09-JUL-1997; 97JP-0183683.
 XX
 PR 09-JUL-1997; 97JP-0183683.
 XX
 (ASAH) ASARI KASEI KOCYO KK.
 DR WPI; 1999-161084/14.
 DR N=PSDB; X14979.
 XX
 PT New slit-like polypeptide - useful for diagnosis and treatment of
 PT nervous, thyroid, adrenal and muscular diseases
 XX
 PS Example 5; Page 27-31; 50pp; Japanese.
 XX
 CC The present sequence represents the full length sequence of a
 CC slit-like polypeptide. The polypeptide is useful for the
 CC diagnosis and the treatment of all nervous diseases, thyroid
 CC diseases, adrenal diseases and muscular diseases.
 XX
 SQ Sequence 1529 AA;
 Query Match 99.44; Score 8265; DB 20; Length 1529;
 Best Local Similarity 99.24; Pred. No. 0;
 Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;
 Qy 1 MRGVGQWMLSLGLVLAIALNKVAPQACGSCSGSTVDCBSGLASVPRNPIRTERL 50
 Db 1 mrgvgqwmllslglvllaialnkvpagqacgscsgstvdchglalrsvprnpiaterl 50
 Qy 61 DLNNGNITRITRTPAGRLRLRLVQLMKNSTIERGAPQDLKELERLRNRHGLQFPE 120
 Db 61 dlngnitrtrtkdfagrlrlrlvqlmknstiergafqdlkelerlrnrhglqfpe 120
 Qy 121 LFLFGTAKLYRLDLSENQIAIPKAFKAGAVIKNLQDNYOISCEGAFALRDLV 180
 Db 121 llflfgtaklyrldlseqnqiaiprkafgaviknlqdyngiscliegafalrdl 180
 Qy 181 TLNNNNITRLSVSNFNMKPLTRFLHNSNYYCDLCHLAWDLKRPVGLTYCGMPS 240
 Db 181 tlennnitrlsvsnfnnmkpltrflhnsnyydcchlawnlksrprvlgtycgmps 240
 Qy 241 HLHGNVAEQKRFVCSDEEKGHGFMAFSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 300
 Db 241 hlhgnvaeqkrfvcs---ghgfnafscsvlhcpaactcsnnivdcrgkglteipt 296
 Qy 301 NLPEITELRIEGLNTIKVIPGAFSPYKRLRIDLSNNISELAPDAGFLRSLNSLVLY 360
 Db 297 nlpeitelrieglntikvipgafspykrlridlsnniseldapdagflrslnslvly 356
 Qy 361 GNKITELPKSLFEGLSIQLLLLNANKINCLRVDAFDLHNNLLSLDYNKLTQANGTF 420
 Db 357 gnkiteplkslfeqlsiqllllnankinclrvaqfdlhnllslsldynkltqangtf 416
 Qy 421 SPLRAIQTMHLAQNPFICDCHLWADLYLHNPETSGARCTSPRLANKRIQKLSKF 480
 Db 417 splraiqtmhlgaqpficdchlwadlylhnnpetsgarctsprlankriqklskf 476

Qy 461 RCS-----GTEDYRSKLSGDCFDALACPEKRCRGTTVDCSNQKLNKIPHEIPQTTA 532
 Db 477 rcsaeyfigpgetdyrsklsqdcfdalacpekrccrgttvdcnqknkipehipqyta 536
 Qy 533 ELRLNNEFTVLEATGFKFKPKLRKRNFSNNKTTDEBGFAGSAGSWELLTSNRLN 592
 Db 537 elrlnneftvleatgfkfkpklrkrfnsnnktdeegafegagsvneilltsnrlen 596
 Qy 593 VORHMKFGLSLKTLMLSRNRTICVNGDSFGLSSVRLSLDNDITTVAGAFDTLHSL 652
 Db 597 vqhkmkfgleslktlmlsrnrticvngdsfglssvrrllsldnditvtagafdtl 656
 Qy 653 STINLANPFWNCNYTLAWLGEMLKRRIVTGNRCQKPYFLKEIPIDQVAIQDTCCDGN 712
 Db 657 stinlanpfnwnytlawlgewlrrkrivtgnrcqkpyflkelplqdvaidqtdcgn 716
 Qy 713 DMSCSPLSRCPTECTCLDTVVRCSNKGVLVPGKIPROVTEVLDGNGFTLVKELSNY 772
 Db 717 dmscsplsrctectcldtvvrscnkgvlvpgkiprovtelydngftlvkelsny 776
 Qy 773 KHLTLIDLNSNRISTLSNQSFSNMWTLTLISNKLRCIPPTFDGLSLLESLHGDND 832
 Db 777 khltildlnsnrilstsngsfnsmtgltilisnrlrcipptfdglsrlslhgdnd 836
 Qy 833 ISVVPFGADNLSALSLSAIGANPLYCDNCNQMWSLSDWSEKPEGIARCAGSGEMADKL 892
 Db 837 isvvpfgadnlsalslsaiiganplycdncnmqwsldwsekeyglarcagsgemadkl 896
 Qy 893 LLTPSKKFKTGCGPVDVNLAKNCPCLSNPKNDGTCNSDVPDFYRCTPGYKQCDV 952
 Db 897 lltpskkfktagcpvdvnlakncpcclsnpkndgtcnsdvpdfyrctpgykqcdv 956
 Qy 953 PIACISNPKHGSGTCHLKGSEEDGFWICADGFEGENCENNVDDCEDNCENNSTCVDG 1012
 Db 957 piacisnpgksgtchlkgeedgfvcicadgfegecnenvddcednccennstcvdg 1016
 Qy 1013 INNYTCLCPPEYTGELCEKILDFCAQDLNCPQHSKCLITPFGKCDCTPGYGEHCID 1072
 Db 1017 innytcclppeytgelceekildfcaqdlncpqhskclitpfgkcdctpgygehcid 1076
 Qy 1073 PDCCDNKKCKNAGACTDANVGTCTCIPCEGSLPCFSPMVLPTSPCDNFCQNGAQC 1132
 Db 1077 pdccdnkcnagactdavnvgctcipcegsylpcfspmvlpstpcdnfcdngaqc 1136
 Qy 1133 IVRINEIQCPLGTQGEKCEKLSVSNFINRESYLTQPSAKVRPTNITLIQATDESGI 1192
 Db 1137 ivrineiqcplgtqgekeckelsvsnfinresylytqpsakvrptnitliqatdesgi 1196
 Qy 1193 LLYKGDKDEIAVELYGRVSRASYDTGSHPASAYSVETINDGNFHIVELLALDGLSLSV 1252
 Db 1197 llykgdkdeiavelygrvsrasydgtshpasaysvetindgnfhivellaldgslsv 1256
 Qy 1253 DGGNPKIITNLSKSTLNFDSPLVYGMGPKSNVSLAQAPQNGTSHFCICRLNLYNSE 1312
 Db 1257 dggnpkiitnlskgtlnfdsplvygmgpksnvslraqapngtshfcicrlnlynse 1316
 Qy 1313 LQDFQKVPWQVIGLPGCEPCHKVCABGTCPQSSQAGTCBCQBGWMLCDQRTNDPCL 1372
 Db 1317 lqdfqkvpmqvgilpgcepchkvcbgtcpqssqagtcbbqbgwmlcdqrdndpcl 1376
 Qy 1373 GNKCVSTGCLPNAFYSYSCKLEHGGVGLDDEEDLFNPOAKIKHGKLSGLSGQYPC 1432
 Db 1377 gnkcvstgclplnafysyscklehgvgvlddeedlfnpocakikhgklsghsgqy 1436
 Qy 1433 ECSSYTGDCSREISRCGERIRDYKQOQYACQCTTKVSRLECGGAGCGCCPLR 1492
 Db 1437 ecssytdgcdreisrcgerirdykqqyacyacttkvslrcggagcgccplr 1496
 Qy 1493 SKRRYSPECTDGSFVDEKVKVKGCTRCVS 1525
 Db 1497 skrryspectdgsfvdevkvkvgctrcvs 1529

RESULT 4
Y27142
ID Y27142 standard; protein; 1503 AA.
XX
AC Y27142;
XX
DT 15-SEP-1999 (first entry)
XX
DE Human slit-2 mature protein (Seq ID No: 6 of JP1164690).
XX
KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;
KW muscle; endocrine system.
XX
OS Homo sapiens.
XX
PN JP1164690-A.
XX
PD 22-JUN-1999.
XX
05-DEC-1997; 97JP-0335435.
XX
PR 05-DEC-1997; 97JP-0335435.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
DR WPI; 1999-411830/35.
DR N-PSDB; N89162.
XX
PT New vertebrate slit protein - useful for diagnosis and treatment of
PT cancers in nerves, muscle and endocrine system
XX
PS Claim 5: Page 43-47; 102pp; Japanese.
XX
CC The invention relates to a vertebrate-derived protein containing an amino
CC acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein
CC has at least 55 % homology to one of sequences shown in Y27141-Y27143,
CC and has slit protein-like activity. The vertebrate slit proteins encoding
CC nucleic acid sequences have at least 60% homology to nucleic acid
CC sequences M89161-163. The vertebrate-derived proteins can be produced
CC recombinantly by transforming host cells with expression vectors
CC comprising the encoding nucleic acids. The proteins of the invention are
CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine
CC system.
XX
SQ Sequence 1503 AA;

Query Match 97.84; Score 8137; DB 20; Length 1503;
Best Local Similarity 99.18; Pred. No. 0;
Matches 1494; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 27 ACAQACSCSGSTVDCHGLALRSVPNIPNTERDLNGNINTRITKDFAGLHRLVQL 86
Db 1 acpaqscscsgstvdchglalrsvpnripnterdlngnintritktfaghlrlvql 60
Qy 87 MENKISTIERGAFQDLKELERLRNHNHQLPPELLPLGTAKLYRLDLSNQIAIPKRA 146
Db 61 menkistiergafqdlkelerlrnahnqilpelflgtaklyrldlsenqialipkra 120
Qy 147 FQGVADIKNLQLDVNYQISCDGAFRALRDLEVLTINNNTIRLSVASFNMHPKIRFLR 206
Db 121 fqgvadiknlqldvnyqiscldgafrralrdlevltinnntrirsvafnmhpkirflr 180
Qy 207 HSNMLYCDCHLAWLSOWLKRKPRVGLYTCQMGPSHLRGNVAENQKREFVCSDEEGBHQS 266
Db 181 hsnmlpdcchlawsdlwlrqprvrglytcmgpsghlrgnvaenqkrefvcs---ghqs 236
Qy 267 FMAPSCSVLHCPAACCTSNVIVDCRGKGLTIPNWLPEITIRLEQNTIKVPPGAFSP 326
Db 237 fmapscsvlhcpaactsnvivrdrqglteipnlpetiteirleqntikvippgafsp 296
Qy 327 YKKLRDLISNQIISLAPDAFGQLRSLSNVLVYGNKITEPKSLFEGFLSLQLLLNAN 386

Db 297 ykkrlridlsnqislapdafgqlrslnsvlvynkitekpslfegflslqlllnna 356
Qy 387 KINCLRVDAFDQLHNLNLSLYDNKLTQAKGTFSPLRAIQTMHLAQNPFTCDCHLAWLA 446
Db 357 kinclrvdafqdlhnlinslydnkltqakgtfslpraigtmhleaqnpfiodchlkwa 416
Qy 447 DYLTNPFTSGARCTSPRLANKRIGQIKSKFKFCS-----GTEYRSKLSDDCPA 498
Db 417 dylnhtnplstsgarctsprlnkrigqlskfkfcsakeqyftptedyrsklsdgcfa 476
Qy 499 DIACPEKRCRGSTVDCSNQKLNKIPHIPOVTAELRLNNEPVLBARCIPKPLQLRK 558
Db 477 diacpekrcregstvdcscngklnkiphiptytaelrlnneftvleatgiklplqirk 536
Qy 559 INFSNKKTIDIEGAFEGASGVNELLTSNRLNVOHKNFKGLSLSLTLMLNSNRITCWG 618
Db 537 infsnkntidieegafegasgvnelltsnrlenvghknfkglslktlmlnsnritycg 596
Qy 619 NDSFVGLSWSRLLSLDNQITTVAGAFDLHSLSTLMLNPPNFCYLAWLGEKLRK 678
Db 597 ndsfvglssvrlslydnqitvagaftdlhslstlmlnppnfcylawlgewlrrk 656
Qy 679 RIVTGPAQKQPFYKLEIPQDAIQDTODGNDNCSPLRCPTCTCLDTWRCNS 738
Db 657 rivtgpaqkqpfykelpiqdvaigdtodgdndnscsplrcptctcldtwrrcns 716
Qy 739 KGLKVLPAIGPFDVETELYLDGNGQTLVPKLSNHYKHLTLDLSNRLSTLSNQSFNMTQ 798
Db 717 kgllkvlpgpfdvetyldngqtlvpklsnhykhltdlidsnrlstlsnqsfnsmtq 776
Qy 799 LLTLLSTNRLRCIPPTFDGLKSLRLSLHGNDISVVPGEAFNDLSHLAIGANPLY 858
Db 777 lltllslynrlicpprtfdglskrlslshgndisvvpgeafndlsahlgaigply 836
Qy 859 CDDNQLWSDNVKSEYKEPGIARCAAGEMADELLTTPSKKFTQGGPVDNVLKNCPC 918
Db 837 cddnqlwdsnvkseykepgiarcagpemdellttpskkftcggpvdnlkncpc 896
Qy 919 LSNPCNDGNTCSNPDVDFYRCTCPYFGKQDQCVIHACISNPKHGKGTCHKEGEEDGF 978
Db 897 lsnpcndgntcsnpdvdfyrcctpyfgkqgdqcvihacisnphgkgtchkegeedgf 956
Qy 979 WCICADGFEBCENGVNDDCENDCENNSTCVGGINNTCLCPETVGLCEKLLDFCAQ 1038
Db 957 wcicadgfebcengvnddcendcennstcvdginnytlclpetyglceeklldfcaq 1016
Qy 1039 DLNFCQDRSKCLLPKFGKCDCTPGYVGEHCDDFDQDQNKKNAGHCTDANGVYTCIC 1098
Db 1017 dlnfcqdrskcllpkfgkcdctpgyvgehcddfdqdnknkngahctdangvtycic 1076
Qy 1099 PEYSGLICEFSPNWLPRTPSCDNFQDQAGCIVRIENIQCLPQYGGKCEKLVNS 1158
Db 1077 peysglicefspnwlprtpscdnfdqagcivrieneicqlpypyggekclvns 1136
Qy 1159 NFINKESTQLPSANVVRQNTWTLQIATDESSGLILYKGDHDAHVELYRGNVSTDTG 1218
Db 1137 nfinkesqlpsavvrqntwtlqiatdesglllykgdhdahvelyrnrvastdtg 1196
Qy 1219 SHPASALYSVETINDGNHIVELLALDQSLSDVGGNPKITLWLSQRTLNFDSPLTWG 1278
Db 1197 shpasalysvetindgnhivellaldqslsdvggnpkitlslsqrtlnfdsppltwg 1256
Qy 1279 GMPGKSNVASLQAGQNGTSGHCRNLYINSELQDQKVPYMGTLPGCEPCHKVCVA 1338
Db 1257 gmpgksnvaslraqngtstghcrlnyinseledqdkvpmgtlpgcepchkvca 1316
Qy 1339 HGTCQPSQAQPTCEQBGWMLPQCDRTNDPLGNKCHVGTCLPINAFYSCKCLEGGH 1398
Db 1317 hgtcqpsaqaptceqbgwmlpdcdrtdnplgnkchvgtclpinafysckcleghg 1376
Qy 1399 GVLCDREELDPNCAQKCKHCKRLSLGLQCPKCSGSGYTGSDCRELSRIGRIIDY 1458

Qy 1399 GVLDEEDLFWQPAKCKKCKRLSGQPVCESSGYTGDSCDREISCRGIRDY 1458

Db 1377 gvldeedlfwqpaickckckrlslgqpvcecssgytgdscdreiscrgirdy 1436

Qy 1459 QKQKQYACQYTKKYSRLCEKGGAGGCGGCLPRLSKRRKYSFECTGSSFVDEKVKVK 1518

Db 1437 qkqkqyaaqcttkvslrlecgaggcgagcgplrlskrrkysfectedgssfvdevkvk 1496

Qy 1519 GCTKCVS 1525

Db 1497 gctcvcs 1503

RESULT 6

Y76117

ID Y76117 standard; Protein: 529 AA.

XX

AC

XX Y76117;

27-MAR-2000 (first entry)

XX

DE Rat Silt homologue, SEQ ID NO:396.

XX

KW Skin: dermal papilla; keratinocyte; neonatal foreskin fibroblast;
 KW embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
 KW secreted; transmembrane; inflammation; cancer; neurological disease;
 KW angiogenesis; tumour vascularisation; growth disorder;
 KW developmental disorder; skin wound; hair follicle disorder;
 KW anti-inflammatory; cytostatic; neuroprotective; vulnary.

XX

OS Rattus sp.

XX

PN W09955865-A1.

XX

PD 04-NOV-1999.

XX

PF 29-APR-1999; 99WO-N20005.

XX

PR 29-APR-1998; 98US-006977.

XX

PR 09-NOV-1998; 98US-018893.

XX

PA (GENE-) GENESIS RES & DEV CORP LTD.

XX

PI Strachan L, Sleeman M, Patson JD, Onrust R, Kumble A, Marison JG;

XX

DR WPI; 2000-072177/06.

XX

N-PSDB; 061825.

XX

PI Novel polynucleotides useful for the treatment of various conditions

XX

PT including wounds and cancer -

XX

PS Claim 4; Page 223-226; 235pp; English.

XX

CC The invention relates to novel nucleic acid sequences derived from rat
 CC dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
 CC and mouse embryonic skin, keratinocyte stem cells and transit amplifying
 CC cells. Polypeptides of the invention may be used to treat inflammation,
 CC cancer and neurological diseases. The proteins may be used to stimulate
 CC the growth and motility of keratinocytes, to inhibit the growth of
 CC cancer cells, to modulate angiogenesis and tumour vascularisation, to
 CC modulate skin inflammation, to modulate epithelial cell growth and to
 CC inhibit binding of HIV-1 to leukocytes. The invention may also be used to
 CC treat growth and developmental defects, skin wounds and hair follicle
 CC disorders. Sequences Y75942-Y76123 represent polypeptides encoded
 CC by cDNA sequences derived from several mouse, rat or human skin cell
 CC types. Sequences Y75942-Y75947, Y76020-Y76021, Y76094-Y76104 and
 CC Y76119 are proteins with an N-terminal signal sequence, indicating that
 CC they are secreted. Sequences Y75986-Y75989, Y76061-Y76071, Y76106-Y76109
 CC and Y76121-Y76122 are proteins with one or more putative transmembrane
 CC domains.

XX

SQ Sequence 1529 AA;

Query Match 97.0%; Score 8065; DB 21; Length 1529;
 Best Local Similarity 96.0%; Pred. No. 0;
 Matches 1471; Conservative 29; Mismatches 21; Indels 12; Gaps 2;

Qy 1 MRGVQWQMSLSLGLVAILNKVAPQAPQACSCSGSTVDCGHALRSVPNIPNTERL 60

Db 1 msvigvqltllslslvlslnkavhaphacpcscsgstvdchglalsrvpnipntrlerl 60

Qy 61 DLNCGNITRITKTPDAGLHRLVQLMKENKSTIERGAFQDKELERLNRNLQLPPE 120

Db 61 dlncgnitritktfdaglrhlrvlqlmkentstiergafqdkelerlrlnrnlqlppe 120

Qy 121 LLFLGTAKLVRLDLSNQIQAIKPKAFRGAVDINKLQDYNQISCTEDGAFRALDLEVL 180

Db 121 llflgtaklvrlldlsnqilqaiakfrgavdinklqldynqisctedgafraaldlevl 180

Qy 181 TLNNNITRLSVASFNMHMKLRTFLHSNNLYCCHLAWLSQWLKRPVGLYTQCGPFS 240

Db 181 tlannnitrlsvasfnhmkrlrtfrlshnnlycdchlawsdlwrqprvlytqcgpfs 240

Qy 241 HLGHNVAQVQKREVFCSDEECHOSSPACSVLHCACTCSNNIVDCRGLTEIPI 300

Db 241 hlghnvaqvqkrvfcs---ghgfsmapscsvlhclactcsnnivdcrglkteipt 296

Qy 301 NLPEITITRLNQWIKYIPPAFSPFKKLRLDLSNQISLHAPDAQGLASLUSLVY 360

Db 297 nlpetitrlngnsrvlppgafspfkklrlldlsnqiselapdaqglrslnslvly 355

Qy 361 GNKITELPSLFEGLFSQQLLLNANKINCLRVADPDLNLLNLSDNKLITARGFT 420

Db 357 gnkitelpslfeqlfsqlllnankinclrvaqdlhnlslldynklgtvaktft 416

Qy 421 SPLRALQTMHLAQPFICDCHLWADLYLTNPISGARCTSPRLNKLRIQIKSKKF 480

Db 417 salralqtmhlqappfidchlwadylhtnpisgarctsprrlnklrqgkskf 476

Qy 481 RCS-----GTEDYRSLKSGCDFADLACEPKRCCEGTVDVCSNOKNLKPHIPIYTA 532

Db 477 rcsakeqfyigtedyrsklsgdcfdadlaacepkrccegtvdvcsnqnlkphipyta 536

Qy 533 ELRLNNEPFTLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGVNELLTSRLN 592

Db 537 elrlnneepftleatgiffklpqlrklnsnkkitdiegafegasgvneilltsrln 596

Qy 593 VQHNKFGLESLATMLRSNRITCVGNSFGISLVSLSLDYNQITVAPGAFITLSL 652

Db 597 vqhnkfgleslatmlrsnritcvngnsfgislvslslldynqitvapgafitlsl 656

Qy 653 STLNLANPFCNLCYLAWLGEAKRKIRVTGNPRCKQPYFLKEIPIQVDAQITCDGN 712

Db 657 stlnlanpfcnchlawlgeakrkirvtgnprckqpyflkeipigvdaqitcdgn 716

Qy 713 DDNSCSPLSRCPTECTCLDVTWRCSNGLVLPKGPIDVTLYLDQNPITVPLKESNY 772

Db 717 ddnsccsplsrpctectcldtvwrscnglvlpkgpidvtyldqnpitvplkcsny 776

Qy 773 KHLTLDLSNRISTLSNQSFNSMQLTLLSYNRLRCPPTFDGLSLRLSLND 832

Db 777 khltdlslsrstlsnqsfnsmtqltllsynrlrcpprtfdglsrlslnd 836

Qy 833 ISVPEGAFNDLSLSLHLAGNPLTCQNMQVLSQWVSEKTEKPEIARCAPGEMAKLI 892

Db 837 isvpegaflndlsllhlagpnlydcnmqvlswvsekepeiarcagpegemakli 896

Qy 893 LLTPSKKFTQGGPVDWNLAKNCPLSNCKMGDCNSGVDFETRCTPYFGQDCDV 952

Db 897 lltpskkftqggpvdwvnlakncplsnckmgdcnsdvdfetrtctpyfgqdcdv 956

Qy 953 PIHACISNCPKSGTCHLKGSEDFWCICADCFEGENCFVNWDCEDDNNNSCTVDG 1012

Db 957 pihacisnckpghgtchlkgeendfwctcadcfegesdnddceddnnnsctvdg 1016

Qy 1013 INNYTCLCPPTGELCEKLDCAQDLPQCHDSKCLITPKGFKCCTPGVGEHCID 1072
 Db 1017 innytklcpptgelceekldfcaqdlpqchdskclitpkgfkcdctpgygehoid 1076

Qy 1073 FDDCCDNKCKNGAHTDAVNGYTCICPGSYGLCFSPFPMVLRTSPCDNFCQNGAQ 1132
 Db 1077 fddccdnkcnghatdavngytcvcpesgylcfesppmvlrtspcdnfcqngaqc 1136

Qy 1133 IVRINEPICQPLGYGCEKELVSVNFINKESYLQISAKVRPQTNTILQIATEDSGI 1192
 Db 1137 iivrnepicqplgylgckelvsfnvfnkesylqipeakvrpqtntilqiatdesgl 1196

Qy 1193 LLYKGDKDIIVELYRGVRASYDTGSPASAIYSVETINDGNFHIWELLALQSLSLV 1252
 Db 1197 llykgdkdliivelyrgvrasydgtshpasaiysvetindgnfhiwelltdslslv 1256

Qy 1253 DGGSPFIITNLKQSTLNFDSPLYVGGMPKSNVASLRQAPQNGTSFHCICINLYNSE 1312
 Db 1257 dggspfiitnlkqstlnfdsplyvggmpknvaslrqapngqtsfhcicirlnlyns 1316

Qy 1313 LQDQFQVMTQIGLPGCECHKKVCAHGTCPQSQAGFTCECQSGWMLCDQRTNDPCL 1372
 Db 1317 lqdqfvmtqiglpqcechkkvcahgtcpssqsgftceceegwmlcdqrtndpcl 1376

Qy 1373 GNKVGHGTCLINAFYSYCKCLBHGGLCDEHDLFNPQAIKCKHGKCLSLGLQPYC 1432
 Db 1377 gnkvghgtclinafysyckclbghgvlcdeedlfnpqavikckhgkclslglgpyc 1436

Qy 1433 ECSSGYTGSCDREISCRGERIDRYQGGYGAACOTTAKVSLBRCGGCAGGCQGPLR 1492
 Db 1437 ecssgytgdscdreiscrgeridryqgggyaacottkvsrlbrcggcaggcqgplr 1496

Qy 1493 SKRKYSFECTDGSFVDEKVKVKGCTRCVS 1525
 Db 1497 skrkysfectdgsfsvdevkvkcgctrcas 1529

RESULT 7

Y9395

ID Y9395 standard; Protein: 1523 AA.

XX

AC Y9395;

XX

DT 08-AUG-2000 (first entry)

XX

DE Human PRO1336 (UW691) amino acid sequence SEQ ID NO:198.

XX

Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX

OS Homo sapiens.

XX

PN W0200012708-A2.

XX

PD 09-MAR-2000.

XX

PF 01-SEP-1999; 99WO-US20111.

XX

PR 01-SEP-1998; 98US-0098716.

PR

PR 01-SEP-1998; 98US-0098749.

PR

PR 01-SEP-1998; 98US-0098750.

PR

PR 02-SEP-1998; 98US-0098803.

PR

PR 02-SEP-1998; 98US-0098821.

PR

PR 02-SEP-1998; 98US-0098843.

PR

PR 08-SEP-1998; 98US-0099536.

PR

PR 09-SEP-1998; 98US-0099596.

PR

PR 09-SEP-1998; 98US-0099602.

PR

PR 09-SEP-1998; 98US-0099642.

PR

PR 10-SEP-1998; 98US-0099741.

PR

PR 10-SEP-1998; 98US-0099754.

PR

PR 10-SEP-1998; 98US-0099763.

PR

PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.
 PR 10-SEP-1998; 98US-0099815.
 PR 10-SEP-1998; 98US-0099816.
 PR 15-SEP-1998; 98US-0100385.
 PR 15-SEP-1998; 98US-0100388.
 PR 15-SEP-1998; 98US-0100390.
 PR 16-SEP-1998; 98US-0100584.
 PR 16-SEP-1998; 98US-0100627.
 PR 16-SEP-1998; 98US-0100661.
 PR 16-SEP-1998; 98US-0100662.
 PR 16-SEP-1998; 98US-0100664.
 PR 17-SEP-1998; 98US-0100683.
 PR 17-SEP-1998; 98US-0100684.
 PR 17-SEP-1998; 98US-0100710.
 PR 17-SEP-1998; 98US-0100711.
 PR 17-SEP-1998; 98US-0100919.
 PR 17-SEP-1998; 98US-0100930.
 PR 18-SEP-1998; 98US-0100848.
 PR 18-SEP-1998; 98US-0100849.
 PR 18-SEP-1998; 98US-0101014.
 PR 18-SEP-1998; 98US-0101068.
 PR 18-SEP-1998; 98US-0101071.
 PR 22-SEP-1998; 98US-0101279.
 PR 23-SEP-1998; 98US-0101471.
 PR 23-SEP-1998; 98US-0101472.
 PR 23-SEP-1998; 98US-0101474.
 PR 23-SEP-1998; 98US-0101475.
 PR 23-SEP-1998; 98US-0101476.
 PR 23-SEP-1998; 98US-0101477.
 PR 23-SEP-1998; 98US-0101479.
 PR 24-SEP-1998; 98US-0101738.
 PR 24-SEP-1998; 98US-0101741.
 PR 24-SEP-1998; 98US-0101743.
 PR 24-SEP-1998; 98US-0101915.
 PR 24-SEP-1998; 98US-0101916.
 PR 29-SEP-1998; 98US-0102207.
 PR 29-SEP-1998; 98US-0102240.
 PR 29-SEP-1998; 98US-0102307.
 PR 29-SEP-1998; 98US-0102330.
 PR 29-SEP-1998; 98US-0102331.
 PR 30-SEP-1998; 98US-0102484.
 PR 30-SEP-1998; 98US-0102487.
 PR 30-SEP-1998; 98US-0102570.
 PR 30-SEP-1998; 98US-0102571.
 PR 01-OCT-1998; 98US-0102684.
 PR 01-OCT-1998; 98US-0102687.
 PR 02-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0103258.
 PR 06-OCT-1998; 98US-0103449.
 PR 07-OCT-1998; 98US-0103314.
 PR 07-OCT-1998; 98US-0103315.
 PR 07-OCT-1998; 98US-0103328.
 PR 07-OCT-1998; 98US-0103395.
 PR 07-OCT-1998; 98US-0103396.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103633.
 PR 08-OCT-1998; 98US-0103678.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.
 PR 26-OCT-1998; 98US-0105694.
 PR 27-OCT-1998; 98US-0105807.
 PR 27-OCT-1998; 98US-0105881.
 PR 27-OCT-1998; 98US-0105882.

PR 27-OCT-1998: 9805-0106052.
 PR 28-OCT-1998: 9805-0106023.
 PR 28-OCT-1998: 9805-0106029.
 PR 28-OCT-1998: 9805-0106030.
 PR 28-OCT-1998: 9805-0106032.
 PR 28-OCT-1998: 9805-0106033.
 PR 28-OCT-1998: 9805-0106178.
 PR 29-OCT-1998: 9805-0106248.
 PR 29-OCT-1998: 9805-0106384.
 PR 29-OCT-1998: 9805-0108500.
 PR 30-OCT-1998: 9805-0106464.
 PR 03-NOV-1998: 9805-0106856.
 PR 03-NOV-1998: 9805-0106902.
 PR 03-NOV-1998: 9805-0106905.
 PR 03-NOV-1998: 9805-0106919.
 PR 03-NOV-1998: 9805-0106932.
 PR 03-NOV-1998: 9805-0106934.
 PR 10-NOV-1998: 9805-0107783.
 PR 17-NOV-1998: 9805-0108775.
 PR 17-NOV-1998: 9805-0108779.
 PR 17-NOV-1998: 9805-0108787.
 PR 17-NOV-1998: 9805-0108788.
 PR 17-NOV-1998: 9805-0108801.
 PR 17-NOV-1998: 9805-0108802.
 PR 17-NOV-1998: 9805-0108806.
 PR 17-NOV-1998: 9805-0108807.
 PR 17-NOV-1998: 9805-0108867.
 PR 17-NOV-1998: 9805-0108925.
 PR 18-NOV-1998: 9805-0108848.
 PR 18-NOV-1998: 9805-0108849.
 PR 18-NOV-1998: 9805-0108850.
 PR 18-NOV-1998: 9805-0108851.
 PR 18-NOV-1998: 9805-0108852.
 PR 18-NOV-1998: 9805-0108858.
 PR 18-NOV-1998: 9805-0108904.
 XX
 XX (GETH) GENENTECH INC.
 XX
 PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CX, Wood WI;
 XX
 DR WPI: 2000-237871/20.
 DR N-PSDB; A37017.
 XX
 PT New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX
 Claim 12; Fig 112; 773pp; English.
 CC A37022 to A37144 encode the new isolated human transmembrane, receptor
 CC or secreted PRO polypeptides given in Y9340 to Y9462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. A37145 to A37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 SQ Sequence 1523 AA;

Query Match 68.8%; Score 5717.5; DB 21; Length 1523;
 Best Local Similarity 66.6%; Pred. No. 1e-301;
 Matches 1017; Conservative 217; Mismatches 277; Indels 15; Gaps 9;

Qy 3 GVGQWMSLSLGLNLVLA-ILNKVAPQACPAQCSGSGSTVDCHEKALSVPRNIPATRLR 60
 Db 7 gyyaavraralalalaslvsppavacptkctcsaasvchglravrprlraeri 66
 Qy 61 DLNKNITRIKTDVAGLRLHVLQWENKSTIERAGFQKLELRLNLNHLNQLFPE 120
 Db 67 dldnndtritmdfaglnrlvhlhedngsvlrgafgdkqlerlrlnknklgvlpe 126

Qy 121 LLFLGTAKLYRLDLSNQIQAPKPRAGVADINKLQDYNQISCIEDGAFRALDLEVL 180
 Db 127 llfgspkltirldlsenqigiprkafgltvdnqlgdnhhsiciedgafraledlell 186
 Qy 181 TLNNKNITRIKTVASFNHMKLRTFLRLSNLYCDHLAWLSDWLKRPVGLTQCMGFS 240
 Db 187 tlannnsiriltvafnmpkirtlrhshlycdchlawlswdlrgrtrvggtlcmwv 246
 Qy 241 HLHGNHVAQWQKRFVCSDEEHQSGHSPMASCVS-LHCPAACCSNNIVDCRGKGLTEI 298
 Db 247 hlrgfnvadvqkgyvcpahs-----eppscnansiscspctcsnnivdcrgkglmei 301
 Qy 299 PTKLPTETIRLEQNTIRVPPGAFSPYKLRRLDLSNNOISELAPDAFQGLRSLNSLV 358
 Db 302 panlpeglveirleqnsikaipagafgykklridisknqsdiepdafqglksitlv 361
 Qy 359 LYGNKITELPSLFEGLFSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLOTIANG 418
 Db 362 lygnkitelagldglvslqlllnankinclrvntfgdlgnlnlslvdynglqtsqg 421
 Qy 419 TSPSLRATQTHLQANPFCDCHEWADYLNHNPIETSGARCTSPRLNAKRIQIISK 478
 Db 422 lfapglsgltlqagpfcvdcchlwadyldgnpietsgarcssprlnakrqsiksk 481
 Qy 479 KFCRCSTEDTVKSLGDCPADLACEKRCSTGVTVDCSNQKLNKIPHEIPQTALRLN 538
 Db 482 kfrcscgedyrsrfsecdmldvcpkrcroegtlvdcnqklvrpshlpeyvtldrlnd 541
 Qy 539 NEPTVLEATGIFKRLPQLARKINFSNNKITIEBGAPEGASGVNELLTSNLENVQHKMF 598
 Db 542 neptvleatgikfklpnlrklnsnknkvevregafgaasvglmnlqetlvbgrv 601
 Qy 599 KGLESKITMLRSNRITCVGNDSFGLSVRLSLYDNQITTVAPGAFDTLASLTANLL 658
 Db 602 rgslgltklmlrsnlscvsnctfaglsrvllslydnrtticipgafvtlslctall 661
 Qy 659 ANPNCNCLAWLWENLKRKRITVGNPCQPKYFLKEIPQDVAIQDPTCDGNDNSCS 718
 Db 662 snpncnclavlgwlrkrirvsgnrcqpfkfeikpigtvaidftc-dgneesqc 720
 Qy 719 PLSCRPECTCLDTVVRCSNKGKLVLPKGI PRDVTLEYLDGQVTVKLSNWKHLTL 778
 Db 721 lpsrcpceptcmctvrcsnkgkrlalprgmkdvtleylqgnhlvtpvrsalrhltli 780
 Qy 779 DLSNNRISTLSNQSFWNTQLTLTLLSNRKLRTIPRTFDGLKSLRLSLHGNDSVPE 838
 Db 781 dlsnnsimlntytfsmshlsltlisylrcipahqnglsrlvrlthgndisvpe 840
 Qy 839 GAFNDLSALSLSLIGANPLYCDGNQWLSVWVSKYKEPGIACAGCEMADKLLTTPS 898
 Db 841 gsfndtalslslgltgnplhcdsrlvsewvskgkpgiicarspemmadrlltpt 900
 Qy 899 KFTQPGVDWVNLAKNPNCLSNPCNNDGTCNSDPVFCPTFGVKGQCDVFIACI 958
 Db 901 hrftqpgvdvnlvaknacsisspcnngtctqdpvlyrcacpysgkqkdvptlnci 960
 Qy 959 SNPKBGGTCHLKEGEDGFWCIADGEGENCENVDNDCENNSCTVDGLNNTYC 1018
 Db 961 gnpqbhgtchldeshdkdfscscplfyegrcelndpdcndcennatcvdglvnyvc 1020
 Qy 1019 LCPFNTGELCEKLDQADQVPCNQSCKLITPFGKCDCTGFTVGKRCIDFDQCDQ 1078
 Db 1021 lcpfntgclcedvchvpevlncqheakcpldkgfscecvpyssgkicetdnddcva 1080
 Qy 1078 KCKKNABCTDVAWNTGTCICPGISGLRCSEFSPWMLPRTSPCNDFQCGAQCIVRINE 1138
 Db 1081 kckrcnqavdntingyctctpgfsfpcchppwmlvltsgdpcdyceqagacivwge 1140
 Qy 1139 PICQCLPGTQCEKSEKLVSVNFINESYQLIPSNVRPQNTITLQIATDEDSGLLYKGD 1198
 Db 1141 ptcrcppgfsprocklitvngfykdsyvelasakvrpqnslvqatdkdngllykdg 1200

Qy 1199 KDHVIAEYVGRVSRASVDTGSHPASAIYSVETINDGNFHVILLALQSLSLSDVGNPK 1258
 Db 1201 nbdlalelygghvlyvdislspprttyvsvetvndgqfhsvelvtlnqtlvldvktgk 1260

Qy 1259 IITNLKQSLTINFDSPLYVGMGPKSNVASLRQAPQNGTSPHGCRILNLSINSLDQFQK 1318
 Db 1261 slgkllqkqpvinsplylgtpstglsalrgtdrlpglgfghgchevrinnelgdfka 1320

Qy 1319 VPMOT-GILGPCPCPKHKVCAHGTQCPSSQAGFTCCQEGWMLPLCDQNTNDPLCNKVC 1377
 Db 1321 lppqslgyspgcksc--tvckhglcrsvckdsvcecrpgtvgplcdqearcdplghrch 1378

Qy 1378 HGTCLPINAFSYSCKLRGSGVLCDREEDLPNCPQAIKCKGKRLSLGLQPYVCESSG 1437
 Db 1379 hgvkatgtg-symckaeegygldcnkndsannacsafkchhgchisdqgepyclqpg 1437

Qy 1438 YTGCDREISCRGERIDYTKQKGYAACQTTKVSRLCRGCGAGGCGPLRSKRKK 1497
 Db 1438 fsgehgqenplcgsvvrevirrkgyasctaskvplmcerggc-gpccqptrskrk 1496

1498 YSECTCGSSFYDEYKVKVCKOCTRC 1523
 1497 yfvgctdgsafveerzhlecglac 1522

RESULT 8
 Y27146

ID Y27146 standard; protein; 1523 AA.

XX

XX Y27146;

XX

DT 15-SEP-1999 (first entry)

XX

DE Human slit-3 protein (Seq ID No: 13 of JP11164690).

XX

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;
 muscle; endocrine system.

XX

OS Homo sapiens.

XX

PN JP11164690-A.

XX

PD 22-JUN-1999.

XX

PF 05-DEC-1997; 97JP-0335435.

XX

PR 05-DEC-1997; 97JP-0335435.

XX

(ASAB) ASABI KASEI KOGYO KK.

XX

UR WPI; 1999-411830/35.

DR N-PSDB; X89163.

XX

PT New vertebrate slit protein - useful for diagnosis and treatment of
 cancers in nerves, muscle and endocrine system

XX

PS Disclosure; Page 64-70; 102pp; Japanese.

XX

CC The invention relates to a vertebrate-derived protein containing an amino
 acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein
 has at least 55 % homology to one of sequences shown in Y27141-Y27143,
 CC and has slit protein-like activity. The vertebrate slit proteins encoding
 CC nucleic acid sequences have at least 60% homology to nucleic acid
 CC sequences X89161-163. The vertebrate-derived proteins can be produced
 CC recombinantly by transforming host cells with expression vectors
 CC comprising the encoding nucleic acids. The proteins of the invention are
 CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine
 CC system.

XX

SO Sequence 1523 AA;

Best Local Similarity 66.6%; Pred. No. 1.5e-301;
 Matches 1017; Conservative 216; Mismatches 278; Indels 15; Gaps 9;

Qy 3 GVGWQMLS-LSLGLVLA-ILNKVAPQAPQCSGSGTVDCGSLALSRPNIPNTERL 60

Db 7 gyaavaralalalalalaslvspppavcpctkctasavchdglravrpgprnaerl 66

Qy 61 DLNGNNTITRTITKTFAGLRHLRLVQLMKNKSTTERGAPDLKLEKRLRLNHNHQLPFE 120

Db 67 dldrnmtrtkmdfaglnrlvhledngvsvieragfdgkqlerlrlnknklylpe 126

Qy 121 LFLFLTKALYRLDLSENQIQAPKPAFGAVDINIKLQYINQISCTEDGAFRLRDLEVL 180

Db 127 lflfgstpkltldisenqigqikpafgritdvnknlgnhnsiciedgafrrldleil 186

Qy 181 TLNNNITRLVSASFNMKPLRTFLHNSNNIKCDCHLWLSDWLRKRPVGLYTCMGPS 240

Db 187 tlmmnsirlvtfnhmpkirtlrlnshlycdhawlsgdlrqrtrvggtlcmavp 246

Qy 241 HLGHNWAVQWKEFYECSDDEEBGQSFMAPSCSV--LHCNPACTCSMNIVDCGKGLREI 298

Db 247 hlrfgnvadqkkeyvcaphs-----epscnansisepscpsctsmnivrckgkime 301

Qy 299 PTMPEPTITRLQNTKIVIPGPAFSYKKLRKIDLSNQISELAPDGLRSLNLSVL 358

Db 302 pmlpegivlelqnsikaiepegatgqkklridisnkqisdapdgaeggkisltslv 361

Qy 359 LYGNKTEPRLSPFEGFLSLQLLLNNKICLDVADPQDLNMLNLSYDNKLQTIARK 418

Db 362 lygnkteleagflgdvslqlllnnknclrvatfgdgnllslgdnkltqiskg 421

Qy 419 TSLPLRAITQTHLAQNFPCDCHLWADYLTNPIETSGACTSPRLANKRIGQISK 478

Db 422 faplgsgqlhlaqpfvcdchklwadyldgdpictesgarcesprlankrsgkisk 481

Qy 479 KFCSCSTEDTRKSLGDCFDALCKEKCRCBGTVDCSNKLNKIPHIPTQTLERLNN 538

Db 482 kfcsgsedyrfsseefmdvcekrceogtvdcsngkvlrplpypvtdrlnd 541

Qy 539 NEFTVLEAGTFKKPLQPKRNKNNKTIIDIEGAFBGASGWEILLTSLNLEWQHMF 598

Db 542 nevsvleatgfkfklpnlrknlnsnkikvegafdgasvqelmlnglqetvgvrf 601

Qy 599 KGLSELTMLNRSNRTPCVGNDSFGLSSVRLSDYDQNTTIVAPGADTLRLSTLMLL 658

Db 602 rglsgklmlrnlsgvndtfaglsrvlslgdnrtitttppgattlvslstall 661

Qy 659 ANFPNCNCLAWLGEKRRKRVITGNPRQKPYFLKEIPQVADIQDTCDDGNDSCS 718

Db 662 snpncnchlawlgkvlrkrivsgprckpfikeipqdvadigdtc-dgneessq 720

Qy 719 PLSCRPTCTLOTVRCNKSGLVLPKGIPOVTELYLDGNQITLVPKLSNYKHLZLI 778

Db 721 lpsrpeqctcmvrcsnkgllrplgmkdvteyleghnlavprelsrlhl 780

Qy 779 DLSNNRISTLSNQSFMNTQLTLILSYNRLKIPPTFDGLSLRLLSHGNDISVYPE 838

Db 781 dlsnnsimlntysfmshstlilsyrlrcpvhafnglrslrvltghsdissvpe 840

Qy 839 GAFNDSLALSHLIGANPLYCDONNMGSDWYSEYKEPGIARCAGGEMADKLLTTPS 898

Db 841 gsfndtlshlaltgplhdcslrlwsewkygkpgiarcsspeppedrilltpt 900

Qy 899 KFTCPQGVVDVNLAKNCPLSNPKNDGTCNSPDVFKYCTPGFGQDCQDVPHACI 958

Db 901 hrfgcpqgvdnvakncalscpknngtctgdpvlyrcaycpvgkdcvptinc 960

Qy 959 SNPKRGSTGKLRGEGEDFWICADGPEGNCENVDNDNDENENSTCVGINNYTC 1018

Db 961 gnpckbggtchlsdhkdgfscscplfgcgcelapdcdendennatcvdginnyvc 1020

Qy 1019 LCPPTETGELCEKLDPCAQNLNCPQDSCKLLTPGFKCDCTPGTVGEBCCDFPDQDQ 1078

Db 1019 lcpptetgclceklldpcaqnlncpqnqdscklltpgfkcdctpgtvgebccdfpdqdq 1078

Db 1021 icppnytgelcdvridhcvpeInlqcheakicpldkgfscecvpgygskicetdnddcca 1080
 Qy 1079 NCKNGARCTDAVNGYTCICPEGYSLGCEPFWMLPRTSPCDNFCQGAQCIIVRINE 1138
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1081 hckrhaqcvdtinytctcpqgfsagfchepppwiltqstpdqyecqgaqlvqge 1140
 Qy 1139 PICQCLPQFGQCEKXLSVNFNINRYSLQIPSAKVRPNTITLIATDEDSGILLYKGD 1198
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1141 ptcrcppfagprcklitvfnvgkdsyvelasakvrgqanislqvtdkngillykgd 1200
 Qy 1199 KDHIAVELYGRVASTDTGSHPASAIYSVETINDCNFIVELALDQLSLSVGGNPK 1258
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1201 hplalelyghvlydsalsppttyvsyvetvndqgfhswelvtleqtlnlvdkgtpk 1260
 Qy 1259 IITNLKQSTLNFDSPLYVGMGPKSNVASLQAPQNCQTSFHCICINLYINSELQDPQK 1318
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1261 sigklqkqepavignsplylgilgtstglsalrgtdrplgfyhgcibevrinneklgdkfa 1320
 Qy 1319 VPMQT-GILGCEPCEHKVCABGTCPQSGAGFTCEQCEGWMGFLCDQRTINDPLGKNVC 1377
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1321 lppqslgvsqgcksc-tvckhglcrsvckdsvvcecrpgwtgplcdqardpqlhzhch 1378
 Qy 1378 HTGCLPINFATSYSCXLEHSGVLDEEDLFWPQCAIKCKHCKRSLGQCPKCRSG 1437
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1379 hqkcvatgt-symckcaeygygdldcnkdsanacsfkhghqchisdqegypclpge 1437
 Qy 1438 YTGSDRCDEISCRGEIRIDYTKQQCYAAQQTKKVSLBRCRGAGCAGCCPLASKRRK 1497
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1438 fsqehqcpqpcqlgvrvvrrqkygascatskplnecrpgc-gpqqcgtprskrr 1496
 Qy 1498 YSBCEDGSSFVDEVEKVKGCOTRC 1523
 :||:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|:||||:|
 Db 1497 yfivqctdgssFveevhlecyclac 1522

RESULT 9

Y04137

ID Y04137 standard; Protein; 1523 AA.

XX

AC Y04137;

XX

DT 15-JUN-1999 (first entry)

XX

DE Human slit 3 protein.

XX

KW Human; slit-like protein; slit 3; slit 1; prevention; treatment;
 KW disease; spinal cord; thyroid gland; ovary; prostate; renal gland;
 KW small intestine; heart; trachea; thymus; lymph node;
 KW muscular system; colon.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide

FT /label= signal

FT 18..1523

FT Protein

FT /label= slit_3

XX

PN JP11075846-A.

XX

PD 23-MAR-1999.

XX

PF 02-SEP-1997; 97JP-0236994.

XX

PR 02-SEP-1997; 97JP-0236994.

XX

PA (ASAH.) ASAH KASEI KOGYO KK.

XX

DR WPI; 1999-257695/22.

XX

DR N-PSDB; K13946.

XX

PT New slit-like polypeptide - useful for prevention and treatment of

XX

PT diseases in spinal cord, thyroid gland, ovary, prostate, renal

PT gland, small intestine, heart, trachea, thymus, lymph node, muscular
 PT system and colon

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 661 pcrqgffllkelplgdvalqdfc-dgneesscqlsprcpcgctmetvvrscnkgiral 719
 Qy 745 PKGIPROVTELYLDGNOPTLVKELSNKYHLTLDLNSNRISTSNQSFNSMTQLLTLL 804
 Db 720 prgmpdvtelegnhltavprelsalrhltldlansismlyntfsmnshltll 779
 Qy 805 SYNRKLCPKPTFDGLKSLRLSLHGNDISVVPKAGFNDLSALSHLAGANFLYCDNMQ 864
 Db 780 syrnrcipvhafaglrslrvltlhyndissvpegsfndltshlaigtphodcsir 839
 Qy 865 WLSWQVSEKTEPGICRACGAGGEMADKLLITPSKRTQCGPVDNLLAKNCPISNPK 924
 Db 840 wlewsakgkpgglarcssepmadrlltthpfrgqcgvdlnlvaknclnsspc 899
 Qy 925 NQGTCSNDPVDFRCTCYGFGQCDVPIRACISPKKGGTCHLKBGEEDGFWICAD 984
 Db 900 nngtctdpvelryroacpysykgdctvptncqpcqhgctchlsdskdgsfscpl 959
 Qy 985 GFGEHGEVWVDCRNDNENSTCVDGINNYTCLCPETGELCEKELDFCAQDLNMQ 1044
 Db 960 gfegrcelnpddcndcennatcvgdinnycpccpypgelcdevidhcvpelnq 1019
 Qy 1045 HDSKCLPLKPGKCDCTPGYGEHCDIDFDQDNKCKAGACTDAMVNGTCTICPYSIG 1104
 Db 1020 hskclpldkgfscvcpysykgldctnddcvahrchhagcvtvtingtctpgfsg 1079
 Qy 1105 LFCFSFPFWLPRSPDNFCQNGAQCIKVINSPICQLPGYGEKCEKLVSNVFNKE 1164
 Db 1080 pfcehggpmvllqtpcdqecqsgagclvqgqetrcpccpfgagprekllvntvfykd 1139
 Qy 1165 SYLQISAKVPAQNTITLQIATDESGILLYKGDKNHIAVELYGRVRSATDGGSPASA 1224
 Db 1140 syvelasakvpaqnsisqvatdkngillykngdnplalelyqghrvlydsiseppt 1199
 Qy 1225 IYSEVETINOGNFHIVELLALQSLSDVGNKPIITLNSKSTNDSFSLVGGWPKS 1284
 Db 1200 ysvetvndgfhsevltnqtnlnvdkgtkslkgkqcpaqlinsplygigtst 1259
 Qy 1285 NVASLRQAQNGTSGFHCIRNLYINSELQDFQVPMQT-GILPCSPCHKRYCAHTGQ 1343
 Db 1260 gisalsrgtdrplgfyhycihevrinnelqdfkalpgelvsqgskc--twkhgler 1317
 Qy 1344 PSSQAQCTFCBCQEGWMLCDQRTNDPCLGNKCVHTGLTINAFSPYCKLBEHGGVLCD 1403
 Db 1318 svekdsvvoecrpgvtgplcdqeardpclghrchhgcvtatg-symkcaegysgdld 1376
 Qy 1404 BEEDLNPQAKICKKHKRLSGLGQPYCESSGYTDCSDREISCRGHERIDTYKQGG 1463
 Db 1377 nkdsnancasckhbgchlsdggqpcgpgfsehgqgeqplgqvrvvrlrkgk 1436
 Qy 1464 YAAQCTTKVSLERCGGAGGCGGLPSKRRKYSPFECTDGSFVDEKVKWGCCTRC 1523
 Db 1437 yascatasvplmeergc-gpccocptrskrykyfqtcdgssfsveewherlycglac 1495

RESULT 11
 Y04136
 ID Y04136 standard; Protein; 1496 AA.
 XX
 AC Y04136;
 XX
 DT 15-JUN-1999 (first entry)
 XX
 DE Human slit 3 mature protein.
 XX
 KW Human; slit-like protein; slit 3; slit 1; prevention; treatment;
 KW disease; spinal cord; thyroid gland; ovary; prostate; renal gland;
 KW small intestine; heart; trachea; thymus; lymph node;
 KW muscular system; colon.
 XX
 OS Homo sapiens.
 XX

PN JP11075846-A.
 XX
 PD 23-MAR-1999.
 XX
 PF 02-SEP-1997; 97JP-0236994.
 XX
 PR 02-SEP-1997; 97JP-0236994.
 XX
 (ASAH) ASAH KASBI KOGYO KK.
 XX
 DR WPI: 1999-257695/22.
 DR N-PSDB: X19946.
 XX
 PT New slit-like polypeptide - useful for prevention and treatment of
 PT diseases in spinal cord, thyroid gland, ovary, prostate, renal
 PT gland, small intestine, heart, trachea, thymus, lymph node, muscular
 PT system and colon
 XX
 PS Claim 1; Page 17-21; 48pp; Japanese.
 XX
 CC The present sequence represents a human slit-like protein designated
 CC slit 3. Slit-like proteins can be used for the prevention and the
 CC treatment of diseases in spinal cord, thyroid gland, ovary, prostate,
 CC renal gland, small intestine, heart, trachea, thymus, lymph node,
 CC muscular system and colon.
 XX
 SQ Sequence 1496 AA;
 Query Match 68.6%; Score 5702.5; DB 20; Length 1496;
 Best Local Similarity 67.2%; Pred. No. 6.6e-301;
 Matches 1008; Conservative 211; Mismatches 268; Indels 13; Gaps 7;
 Qy 27 ACQAQCSGSGTVCHGALRLSPVNPINTEBLDNGNITRTKTPDAGLRHLVLQL 86
 Db 6 acptktsaasvchqglravrpgipraeadridnnritkndfaglnrvhl 65
 Qy 87 MENKSTIERGAFQDLKELERLNNHQLFPELLFLGTAKTLRLDLSNOQIAPKRA 146
 Db 66 edogvslervagfdqlqlerlrlnknlqvlfpqstgklrldlsnqgdlpkr 125
 Qy 147 PRGAVDINKLQDYNQISIEDGAFRALRDLVLTNNNNITRLSVASFNPKMLRFL 206
 Db 126 frigtvknlgldnnhisciedgafrairdleltlannnisrlvtsfnhmpkrlrl 185
 Qy 207 HSNKLYCDCHLAWLSDMLKRPVGLVYTCQMGPSLGHNRVAZVQKREFVCSDEEBGQS 266
 Db 186 hsnhlycdchlawlsdwlrgtrvggftlcmaphlrgfnvadqkvyqvcaphs--- 241
 Qy 267 FMAPSCSV-LHCACTCSNNIVDCRGKGLTEIPTNLPEITTEIRLEQNTKIVPPGAF 324
 Db 242 -eppscnansicpspctcsnnivdcrgylmepalnpesivlerleqnskalpaf 300
 Qy 325 SPYKRLRLDLSNNOISELAPAFQGLSLNSLVYGNKITELPKSLFGLFSLQLLNL 384
 Db 301 tpykklridlskqslsdapafgglkslsvlygnkitelakglfslvslqlln 360
 Qy 385 ANKINCLVAPQDLNHLNLLSLYDNKLTQKTGFTSPSLRAIQZMHLAQNPPICDCHLW 444
 Db 361 ankincrvlffqdlqnlslslydnkltqktgftspslraiqzmlhaqnpfvcdchlw 420
 Qy 445 LADYHNPITSGARCTSPRLANKRQIQKSKKPCSGSEDRSKLSGDCPALACE 504
 Db 441 ladyhnpitsgarctsprrlankrqlsqskkrcsgsedrsklsfscfndlvcpce 480
 Qy 505 KRCCEGTIVCSNGLKNIPEIPQYTAELRNALNEFTVLAETGFKPLQLKRNFSN 564
 Db 481 krccegtivcsnqlrvlshpvytdlrndnevsleatgfkfklpnlkrlnsn 540
 Qy 565 KITIDIEGAFSGASWELLISNRLNVOHNMFKGLSIAKTLNRSNRTICVNGDSF 624
 Db 541 kikevregafsgaswqlnltqgltvhrvfrglsglktlmlrnlsgvcsndfag 600

QY 625 LSSVRLSLSDNQITTVAPGAFPTLRLSLTLNLLANPFCNCLAWLGEMLKRRKIVTGN 684
 DB 601 LSSVRLSLSDNQITTVAPGAFPTLRLSLTLNLLANPFCNCLAWLGEMLKRRKIVTGN 660

QY 685 PRQKPFYKLPKIPDQVAIQDPTCDGNDNSCSPLSRCTPCTCLDTVVRCSNKLKVL 744
 DB 661 PRQKPFYKLPKIPDQVAIQDPTCDGNDNSCSPLSRCTPCTCLDTVVRCSNKLKVL 719

QY 745 PKGIPROVTEVLDNQPTLVFKELSNKRLTLDLNNRISTLSNQSFSNNQTLTLIL 804
 DB 720 PKGIPROVTEVLDNQPTLVFKELSNKRLTLDLNNRISTLSNQSFSNNQTLTLIL 779

QY 805 STNKLRCIPPTFDGLKSLRLSLHSGNDISVPEGAFNDLSALSHAIAGNPLYCDNMQ 864
 DB 780 STNKLRCIPPTFDGLKSLRLSLHSGNDISVPEGAFNDLSALSHAIAGNPLYCDNMQ 839

QY 865 WLSQWVKSIEKPGIARCAAGGEMADKLLITPFSKKTCCGPDVNNILAKNCPILSNPK 924
 DB 840 WLSQWVKSIEKPGIARCAAGGEMADKLLITPFSKKTCCGPDVNNILAKNCPILSNPK 899

QY 925 GDNKNSDVPDYRCTCPYKGGQDQVPHACTISNPKHGCTCHLAGEEDGFWICAD 984
 DB 900 GDNKNSDVPDYRCTCPYKGGQDQVPHACTISNPKHGCTCHLAGEEDGFWICAD 959

QY 985 STNKLRCIPPTFDGLKSLRLSLHSGNDISVPEGAFNDLSALSHAIAGNPLYCDNMQ 1044
 DB 960 STNKLRCIPPTFDGLKSLRLSLHSGNDISVPEGAFNDLSALSHAIAGNPLYCDNMQ 1019

QY 1045 WLSQWVKSIEKPGIARCAAGGEMADKLLITPFSKKTCCGPDVNNILAKNCPILSNPK 1104
 DB 1020 WLSQWVKSIEKPGIARCAAGGEMADKLLITPFSKKTCCGPDVNNILAKNCPILSNPK 1079

QY 1105 LFCFSPSPWLPRTSPCDNFQDQCAQVIRINPEIQCLPAGTQGEKRLVSVNFMKE 1164
 DB 1080 LFCFSPSPWLPRTSPCDNFQDQCAQVIRINPEIQCLPAGTQGEKRLVSVNFMKE 1139

QY 1165 SYLQIPASVAPQWITLIQIATDDESGILLKGDKHIAVELTGRVRASTDTGSHSPA 1224
 DB 1140 SYLQIPASVAPQWITLIQIATDDESGILLKGDKHIAVELTGRVRASTDTGSHSPA 1199

QY 1225 ITSVETVNDGFWHIVALLQSLSLVDGNGKIITLNSKQSTLNFDSPLVGGMPKS 1284
 DB 1200 ITSVETVNDGFWHIVALLQSLSLVDGNGKIITLNSKQSTLNFDSPLVGGMPKS 1259

QY 1285 NVASLQAPQNGSTFSGRINRLINSELQDQFQWQMT-GILPCEPGRKVKCAHGTQ 1343
 DB 1260 NVASLQAPQNGSTFSGRINRLINSELQDQFQWQMT-GILPCEPGRKVKCAHGTQ 1317

QY 1344 PSSQAGPTCECBQWNGPLCDQDNPCLGNKCYHGTCLINAFSTSCKLCBHGHWLGD 1403
 DB 1318 PSSQAGPTCECBQWNGPLCDQDNPCLGNKCYHGTCLINAFSTSCKLCBHGHWLGD 1376

QY 1404 REEDFLNPQAKNKHCKRLSLGLQPSYCESSGYTDCSDREISCRGERIDRYKQKG 1463
 DB 1377 REEDFLNPQAKNKHCKRLSLGLQPSYCESSGYTDCSDREISCRGERIDRYKQKG 1436

QY 1464 YAAQCTTKVSKLBCRGCGAGGCGPLSRKRRKYSPFECTDGSFVDEKVKMGCTRC 1523
 DB 1437 YAAQCTTKVSKLBCRGCGAGGCGPLSRKRRKYSPFECTDGSFVDEKVKMGCTRC 1495

RESULT 12

Y14142
 ID Y14142 standard; Protein; 1523 AA.

XX
 AC Y14142;
 XX
 DT 22-JUL-1999 (first entry)
 XX
 DE Human Slit protein sequence.
 XX
 KW Slit; human; diagnosis.
 XX

OS Homo sapiens.

XX

PN

W09923219-AL.

XX

PD

14-MAY-1999.

XX

PF

29-OCT-1998; 98WO-US22845.

XX

PR

13-AUG-1998; 98US-0096420.

PR

31-OCT-1997; 97US-0063946.

XX

PA

(OSIR-) OSIRIS THERAPEUTICS INC.

XX

PI

Connolly 7, Rajput B;

XX

DR

WPI; 1999-337485/28.

DR

N-PSDB; X61026.

XX

PT

New human slit polypeptide and polynucleotide

XX

PS

Claim 12; Fig 2; 90pp; English.

XX

CC

This sequence is the human slit polypeptide of the invention.

CC The slit protein is useful for patients in need of the slit polypeptide,

CC and its antagonist is useful for patients in need of inhibition of

CC the slit protein. Diagnosis of a disease or susceptibility to a

CC disease is achieved by determining the presence of a mutation in the slit

CC coding sequence, or determining the presence of slit by detecting

CC expression levels of the slit protein. Anti-slit antibodies are useful

CC for diagnosis of conditions associated with slit protein.

XX

SQ

Sequence 1523 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match

67.5%; Score 5611.5; DB 20; Length 1523;

Best Local Similarity 65.7%; Pred. No. 5.7e-296;

Matches 1002; Conservative 216; Mismatches 293; Indels 15; Gaps 9;

QY 3 GVGWMLSLGSLVLA-TLNVAPQAPCAQCSGSGTDCRGIALSRVPPNIPNTERLE 60

DB 7 GVGAVARIALALASVSGPVEAPCTCKCSAASVCHGILRVPRGIPNAERLE 66

QY 61 DLNNGNITRTITDFAGLRLVQLMENKSTIERGAPQDLKELERLNRNHLQFPPE 120

DB 67 dldntrntitkmdfeglknlrvlhledngvsviergafgdlqglerlknknlqvp 126

QY 121 LFLFGTAKLYRLDLSNQIAPKAPRAVGDVNLQNDYQWISCTIEDGAFRALRDLV 180

DB 127 llfggtpklrtldlsengilgprkafgrtdvknlgldnnhisciedgafalrdle 186

QY 181 TLNNNNITRLVSANFNNMPLRFLRSLNNYLDCCHLAWDLKRLPRVGLTQCMQPS 240

DB 187 tlannnniriltvsnfnnmklrtlrslnnhlydcchlawsdlrgrtrvggtlcmq 246

QY 241 HLRGHNVAQVAKRFVSCDEEGHQSFAVSCFVS-LHCPAACSNNVDCRGHGLTEI 298

DB 247 hlrgfnvadvgkkyvcpaphs-----eppscnansicpsectscnldvrglme 301

QY 299 PTNLPEITRELEQNTIKVPPGAFSPYKRLRLDLSNNOISELAPDQGLRSLSLV 358

DB 302 pmlpeglvtrleqnsikaipagftgkykklrlsdngisdiagdgaqgiksltslv 361

QY 359 LYGNKITLPLSLPGLFLSLQLLLNNAKINCLRVDAFDHNLNLSLSDYDKLIQTAG 418

DB 362 lygnkitlakglfdglvslqlllnnknclrvwtfdglnlnlsldydklqtag 421

QY 419 TFSPLRAITQWHLQNPFTCDCHLWADLYHTNPITSGACSTSPRIANRIGQIKSK 478

DB 422 lfaplgslgtlhnqpfvcdchlwadlyhtnpitsgacstspriankrigrqiksk 481

QY 479 KFCSCGTDYSKSLSGDPCADLACPEKRCRGTVDCSNQKLNKIPPHIQTABELNKN 538

DB 482 kfcscgtdysrfsfscdclvpekercregtivdcsnqklnvrlphlpeytdrlind 541

Qy 539 NEPTVLEATGIFKKLPQLRKINFSNNKITDIEBAGFEGASGVNEILLTSNRLEWQHKMF 598
 Db 542 nevsrleatgikfklpqlrkinlsnnkikeregafdgasvqelnltngletvbgvrf 601

Qy 599 KGSLSLKTLMLNRIITCVGNDSPIGLSVRLSLYDQNTITVAPGAFDTLHSLSTLL 658
 Db 602 tglsglktlmlnrlisniscvndtfaglsrvllaydnrtitttppgtatvslvphhkp 661

Qy 659 ANPFCNCYLAWGLNKKRIVTGNPRQKPYFLKEIPQIDVAIQDPTCDGNDGNSCS 718
 Db 662 vqplqglplawglkwlrrkrivgnprqkpfelkelpqdvaiqnftc-dqneesscq 720

Qy 719 PLSCRPTCTCTDITVWCKNGLKVLKPGIDPDTVELYLDGQFTLVKPELSNWKHLTL 778
 Db 721 lpsrceqctmetvrcsnkglralsprgmpkdvtelylegahlavprelsalrhltii 780

Qy 779 DLSNKRISTLSNQSFSNMTQLLTLLSYNRLACIPPTFDGLSKLRLSLAGNDISVPE 838
 Db 781 dlsnksismltuytfnmshtllslynlrlclpwhafnglrsvlthlqgdissvpe 840

Qy 839 GAFNDLSALSHALGANPLVCDNKNWLDVWVKEKPGIARCAKGEMADLLITTPS 898
 Db 841 gsfndltslshalgatphdcslrlwsewkyagkegiarccspegmadllittpt 900

Qy 899 KKFCTGQPPVWNLAKNCLSNPKNDGTCSNDVDFVYRCTCPYGFQDCDVPFHACI 958
 Db 901 hrfaqgspdvslvakcnalsscpknngtctqdvlyrcacopykgtcdtvpintci 960

Qy 959 SNPKCHGTCHLKEGEEDGFWICADGPGKCNENVDCCDNCENNSCTVDCINMYTC 1018
 Db 961 qncqhgqthldshkdgfscscplgfggcrclngddceendceonatcdvlgimvc 1020

Qy 1019 LCPPEPTGELCEKZKLOFCAQLNPFQHDSCILLPGFKCOTCTPYGHEHCDIDFDQCD 1078
 Db 1021 lcppeptgclcdvhdvcpelnlqcheakclpdkgfsceccpgysgklcetdnddca 1080

Qy 1079 KCKNKAHCTDAWNTGTCICPEYSGLPCEFSPPMVLPTSPCNDFOQNGACIVRINE 1138
 Db 1081 kckrhqacvcdttingyctcpgtgsfpclheppmvlgtscpcdgycqagqivqge 1140

Qy 1139 PTCQCLPTGTCGKELKYSVNFINKESTYLQIPASVPQNTITLQIATDEDSGILLYKGD 1198
 Db 1141 ptcrcpgfapgtcklitvfnvgkdsyvelasakvrqpanislqvatdkngillykd 1200

Qy 1199 KQIHAVELYKGRVASTDGTGSPASAIYSVETINGNFIHVELLADQLSLSDVGNPK 1258
 Db 1201 ndplapelyghvrlvydsvsppttysvetvndggfshsvewtlinglilvdygtpk 1260

Qy 1259 IYTLKSHQSTLMDSPLYVGMPKSNVASLQAQNGQNTSPHGCIRNLINSELDFQPK 1318
 Db 1261 slgftqgcpavinsplylgipstglsalrqgtcdrlpglghgchevrlnnelgftka 1320

Qy 1319 VMQOT-GILGRCPECHKVKCAHGTCPSSAAGFTCEQSGWGLCDORTDNLGNKCV 1377
 Db 1321 lppsglsvsggksc-tvckghlcrsvkdsfvcecrpwtgplcdqearcdpqlgrch 1378

Qy 1378 WTCPLNFAFSKCLGEGHGVLCDEBDFNPQAIKCKHKGKRLGSLGQYFCBCSSG 1437
 Db 1379 hqkvatgt-symckcaegygdldcnkdsanacsafckhbgchldsgqepylcpg 1437

Qy 1438 YTGDSCHREISCGERIDYTYNQQGYACQOTTKVSRLEKGGGAGGCGPPLRSEKRR 1497
 Db 1438 fsgehcqampclqgvvrevirrkgyasctaskvpimecrgg-gpqcqptrskrrk 1496

Qy 1498 YSPECTGDSFVDEKVKVSGCTRC 1523
 Db 1497 yvfqctdgsfveevrhlceqlac 1522

RESULT 13
 W46966
 ID W46966 standard; Protein; 1534 AA.

XX
 AC W46966;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Amino acid sequence of a human slit-like polypeptide.
 KW Slit-like protein; human; diagnosis; treatment; brain-specific disease;
 KW cancer; antibody.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 11..26
 FT /note= "signal peptide"
 FT Protein 27..1534
 FT /note= "mature protein"
 XX
 PN JP10087599-A.
 XX
 PD 07-APR-1998.
 XX
 PP 15-JUL-1997; 97JP-0205351.
 XX
 PR 16-JUL-1996; 96JP-0186219.
 XX
 PA (ASAP) ASAMI KASEI KOYO KK.
 DR WPI; 1998-267127/24.
 DR N-PSDB; V16978.
 XX
 PT Human Slit-like protein - useful for diagnosis and treatment of
 PT brain-specific diseases and cancers
 XX
 PS Disclosure; Pages 31-35; 45pp; Japanese.
 XX
 CC The present sequence represents a novel human slit-like protein (the
 CC mature protein is claimed in Claim 1). The slit-like polypeptide is
 CC useful for diagnosis and treatment of brain-specific diseases and
 CC cancers. Antibodies directed against the protein, or its fragments
 CC can also be used for diagnosing cancer.
 XX
 SQ Sequence 1534 AA;

Query Match 67.34; Score 5597; DB 19; Length 1534;
 Best Local Similarity 65.34; Pred. No. 3.5e-295;
 Matches 992; Conservative 228; Mismatches 283; Indels 16; Gaps 3

Qy 15 LVIALNLPVAPQCAQPCQSCSGSTVDCGHLARLSPVFNIPNTERLDLNGNNITRTKD 74
 Db 21 llvaawrlgssacpalcitctgtvdcvgtglqapknprterlelgnnitrknd 80

Qy 75 FAGLRLELVLYLNNKSTIERGAFGLKKEKLERLNRNHQLPPELLFGLTYAKLYLDL 134
 Db 81 faglkrlvrlmqnlgavergfdmkelrlrlnnqlhmlpellfqnqalsldl 140

Qy 135 SENQIQALPFRAGVADINKLQLYNYSICEDGAFALRQLEVLTLNNNITRLSVAS 194
 Db 141 senaigpalkpfragtdlknrlkngiciseegfralrglevltlmmnntiprsv 200

Qy 195 PNMFLPLRFLNLSNLCCHLAWLSDLMLKRPVGLVLTQCMGPSHLRGNHVAQVQRE 254
 Db 201 pnmflplrlflnlsnlnclchlawlsdlmlkrrpvlglvltqcmgpshlrgnhvavqre 260

Qy 255 FVCSDEBEGGSFMAFGSCVL--HPCAATCSNVIDCRGKGLTEIPLNPETIEALE 312
 Db 261 fcsdgsegag--vptctlsagscpantcsngvdcrgkltaipalnpetmetle 317

Qy 313 QNTIKVIVPGAFSPYKLRRLDLSNMQISELAPDQAFGLRSLNLYGNKITELPKSLF 372
 Db 318 lngikslvpgafspyrklrrldlsnnqlaelapdaggrlslnslvgnkitdprgfv 377

Qy 195 FNEHMKLRTFLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMPGSHLRGHNVAEVQKRE 254
|||||.....|..|||...|||..|..|..|..|..|..|..|..|..|..|..|..|..|..|..|

Query Match : 67.2%; Score 5589; DB 20; Length 1534;
Best Local Similarity 65.2%; Pred. No. 9.5e-295;
Matches 991; Conservative 228; Mismatches 284; Indels 16; Gaps 5;

Search completed: January 22, 2001, 12:13:45
Job time: 1382 sec


```

Query Match          68.94; Score 57.08; DB 2; Length 1523;
Best Local Similarity 66.94; Pred. No: 0;
Matches 1015; Conservative 223; Mismatches 265; Indels 14; Gaps 9;

QY 11 LSLGLVLA-LINKVAPAPQACQSCGSSVDCGHLALASVPNNIPNTERLDLGNMNI 69
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 16 LALALALASTLSGPPAAACPTCKTCSAASVDGCGHLGVAFPGIPFARNERLDLRNIT 75

QY 70 IKTDFAGLRHLNVLQNMENKISTIERGAFQDLKLERLRLNNHLLQFPELLFLGTAKL 129
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 76 ITKDFPTGLKRLHVLQEDQVSVERGAFQDLKLERLRLNNKQLVPELLQSTPKL 135

QY 130 YRLDLSNQIQAIPRAFAGVDIKNLQLDYNGISCTEDGAFRALROLEVLTLNMMNITR 189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 136 YRLDLSNQIIGIPRAFAGVGWGVKNQLDNNHISCTEDGAFRALROLEITLNNMINSR 195

QY 190 LSVASFNMFKPLRTLRRLHNNYLDCCHLAWLSDWKRLSPRVGLTITQMGPSHLRGNVAE 249
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 195 LVTVSFNMPIKRLTLRLHNNYLDCCHLAWLSDWKRLTGGTIGITLQMAPVHLRGFSVAD 255

QY 250 VQKREFVCSDEEGHGSFPAQSCSV--LHCPAACTCSNNVDRCNGKLTETIPTLPEIT 307
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 256 VQKKEVT---PGPSS-PAPCAHNSLSCPSCACSNNVDRCNGKLTETIPANLPGIV 310

QY 308 EIRLEQNTIKTIPPGAFSPYKKLRRLDNNQISLAPADAPQGLSLNSLVYGNKITE 367
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 311 EIRLEQNSIKIPAGAFIQYKKLRRLDNNQISLAPADAPQGLSLNSLVYGNKITE 370

QY 368 PKSLPEGLFSIQLLLNANKNICLRVADQFDLNNLLSLDYNNKLTITAGKTSPLRAIQ 427
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 371 PKGLPGDLVSLQLLLNANKNICLRVTFDQDNNLLSLDYNNKLTITKSLGAPLQSTO 430

```


Qy 428 TMLHAGNFPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFCSGSD 487
 Db 431 TMLHAGNFPCDCHLWADYLTNPITSGARCTSPRLANKRISQIKSKFCSGSD 490

Qy 488 YRSKLSDGCFADLACPEKCRBGTIVDCSNQKLNKIPHIPOYTAEIRLNNNEFTVLEAT 547
 Db 491 YRMSFSSCFADLACPEKCRBGTIVDCSNQKLSRIPSHLPETDRLANDNDVLEAT 550

Qy 548 GIFKLPQLRKINFSNNKITIIEGAFEGAGSCVNEILLNRLNENQKMFGLSLKTL 607
 Db 551 GIFKLPQLRKINLSNNRIKIEVREGAFDAGVQELMLNQLNLETHGKMFGLSLKTL 610

Qy 608 MLRSNRTICVNDSPFGLSVRLSLYDQITTVAPAGDTLHSLNLLANPNWNCY 667
 Db 611 MLRSNLSICVNDSPFGLSVRLSLYDQITTVAPAGDTLHSLNLLANPNWNCY 670

Qy 668 LAMGLNKKRIRVSNPQKPPFLKEPIQDAVQDPTCDGNDGNSCFLSKCPTEC 727
 Db 671 LAMGLNKKRIRVSNPQKPPFLKEPIQDAVQDPTC-BGNHNSCQLSPFQEC 729

Qy 728 TCIQTVRCSNGLIPLKGIPODVTLYLQDQNTVLPKLSNKKHLLIDLSNRRIST 787
 Db 730 TCIQTVRCSNGLIPLKGIPODVTLYLQDQNTVLPKLSNKKHLLIDLSNRRIST 789

Qy 788 LMSGNRTICVNDSPFGLSVRLSLYDQITTVAPAGDTLHSLNLLANPNWNCY 847
 Db 790 LMSGNRTICVNDSPFGLSVRLSLYDQITTVAPAGDTLHSLNLLANPNWNCY 849

Qy 848 SHALGAINPLDCCDGLNLSNKKHLLIDLSNRRIST 907
 Db 850 SHALGAINPLDCCDGLNLSNKKHLLIDLSNRRIST 909

Qy 908 DVMILAKNCPCLNCKNDGTCNSDPVFIKCTCPYGFQKQDQVPHACISNPKHEGT 967
 Db 910 DVMILAKNCPCLNCKNDGTCNSDPVFIKCTCPYGFQKQDQVPHACISNPKHEGT 969

Qy 968 CHLKEEDGSPWCLDGFEGECNVDVDCEDNENNSCTVDGINNTICLCPPTTGE 1027
 Db 970 CHLKEEDGSPWCLDGFEGECNVDVDCEDNENNSCTVDGINNTICLCPPTTGE 1029

Qy 1028 LCEKLPQDQALNCKNDGTCNSDPVFIKCTCPYGFQKQDQVPHACISNPKHEGT 1087
 Db 1030 LCEKLPQDQALNCKNDGTCNSDPVFIKCTCPYGFQKQDQVPHACISNPKHEGT 1089

Qy 1088 TDAVNGTICICPQSGFLCEPSPMVLPRTPSCDNFDQNGQAQICVINEPIQCLQY 1147
 Db 1090 TDAVNGTICICPQSGFLCEPSPMVLPRTPSCDNFDQNGQAQICVINEPIQCLQY 1149

Qy 1148 QGKCEKLVSNVFNKESITLQIPSAKVRPQNTIQTATDEDSGILLYNGKDHIAVELY 1207
 Db 1150 QGKCEKLVSNVFNKESITLQIPSAKVRPQNTIQTATDEDSGILLYNGKDHIAVELY 1209

Qy 1208 RGRVRASTDSHPASIVSETINDGNHIVELLADQSLSLYDGNPKIITNLSKQS 1267
 Db 1210 RGRVRASTDSHPASIVSETINDGNHIVELLADQSLSLYDGNPKIITNLSKQS 1269

Qy 1268 TLMFDSPLIVGMPKSNVSLAQPGQNGTSHGCRINVLINSELQDFQKVRPMT-GIL 1326
 Db 1270 TLMFDSPLIVGMPKSNVSLAQPGQNGTSHGCRINVLINSELQDFQKVRPMT-GIL 1329

Qy 1327 PGCEPCHXVCAHGTQPSQAGTCEQCBGWNGLDQRTINDPCLNCKVBTCLPINA 1386
 Db 1330 PGCEPCHXVCAHGTQPSQAGTCEQCBGWNGLDQRTINDPCLNCKVBTCLPINA 1386

Qy 1387 FYSYCKLBSHGGLVDEEDLNFQAIKHKGRCLSLGQPCYCSGSDGSDSCRE 1446
 Db 1389 FYSYCKLBSHGGLVDEEDLNFQAIKHKGRCLSLGQPCYCSGSDGSDSCRE 1446

Qy 1447 ISCHGERIDYQQQYQAQCTTKYRSLRDCRCSGAGGCGCPFLSKRRKYSEFCTDS 1506
 Db 1447 ISCHGERIDYQQQYQAQCTTKYRSLRDCRCSGAGGCGCPFLSKRRKYSEFCTDS 1505

Qy 1507 SFVEDEVKVGCGTCR 1523

Db 1505 SFVEDEVKVGCGTCR 1522

RESULT 2

742218

slit-1 protein homolog - rat

N:Alternate names: MEHF4 protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 03-Dec-1999 sequence_revision 03-Dec-1999 #text_change 21-Jul-2000

C:Accession: J42218

R:Nakayama, M., Nakajima, D., Nagase, T., Nomura, N., Seki, N., Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motif

A:Reference number: J42218; MIM:98360089

A:Accession: J42218

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1531 <R>

A:Cross/References: EMBL:AB011530; NID:g3449289; PID:BAJ3460.1; PID:g3449290

A:Experimental source: strain Sprague-Dawley; brain

C:Genetics:

A:Gene: MEHF4

Query Match 67.14; Score 5578.5; DB 2; Length 1531;
 Best Local Similarity 65.04; Pred. No. 0;
 Matches 988; Conservative 230; Mismatches 282; Indels 19; Gaps 6;

Qy 15 LVLAILNKPACPAQCSGSCSTVDCHGLALRSVPNIPNTERLQDNGNNITRTITD 74

Db 21 LMLAANLKGATGATPACTCTGTCTVDCGTLQAIKPNPNTRELELNNITRTITD 80

Qy 75 FAGLRHLVQLMKNKSTIERGAFQDKELERLNRNHLQPELLFGLAKLYLID 134

Db 81 FAGLRHLVQLMKNKSTIERGAFQDKELERLNRNHLQPELLFGLAKLYLID 140

Qy 135 SENIQATIPKAFRAGVADIKNLQDQNTVLPKLSNKKHLLIDLSNRRIST 194

Db 141 SENIQATIPKAFRAGVADIKNLQDQNTVLPKLSNKKHLLIDLSNRRIST 200

Qy 195 FNMFKPLTFLBSNHLPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC 254

Db 201 FNMFKPLTFLBSNHLPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC 260

Qy 255 FVCSDEEHSQFMAPSCSVL-HCPAATCTSNIVDCRGLNLEIPLNLPETITRE 312

Db 261 FVCSDEEHSQFMAPSCSVL-HCPAATCTSNIVDCRGLNLEIPLNLPETITRE 317

Qy 313 QNTIKVTPPGAFSPYKLRIDLNSQWIEIAPDQGLRSLNLSLYGNKLTPLKSLF 372

Db 318 QNTIKVTPPGAFSPYKLRIDLNSQWIEIAPDQGLRSLNLSLYGNKLTPLKSLF 377

Qy 373 BELFSLQLLNNANKINCLVDAFQDLNLLSLYDGNPKIITNLSKQS 432

Db 378 BELFSLQLLNNANKINCLVDAFQDLNLLSLYDGNPKIITNLSKQS 437

Qy 433 QNFPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC-----G 484

Db 438 QNFPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC-----G 497

Qy 485 YRSKLSDGCFADLACPEKCRBGTIVDCSNQKLNKIPHIPOYTAEIRLNNNEFTV 544

Db 490 YRSKLSDGCFADLACPEKCRBGTIVDCSNQKLSRIPSHLPETDRLANDNDVLE 555

Qy 545 BATGTFKLPQLRKINFSNNKITIIEGAFEGAGSCVNEILLNRLNENQKMFGLSL 604

Db 556 BATGTFKLPQLRKINLSNNRIKIEVREGAFDAGVQELMLNQLNLETHGKMFGL 615

Qy 605 TMLHAGNFPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC-----G 664

Db 616 TMLHAGNFPCDCHLWADYLTNPITSGARCTSPRLANKRIGQIKSKFC-----G 675

Qy 665 NCTLANLWGLKKRKIVTGNPRQKPYFLKEIPIDQVAIQDPTCDGNDNNSCSPLSRCP 724
 Db 676 NQCLANLWGLKKRKIVTGNPRQKPYFLKEIPIDQVAIQDPTCDGNEGVCGLSRPQCP 735

Qy 725 TECTCTDITVRCSSNKGKLVKPGIDPVTELYLDGNQFTVLPKLSNKHLLTDLSSNR 784
 Db 736 QECACDITVRCSSNKHLLQALPKGIPKNTVTELYLDGNQFTVLPQSLTFTYQLQVLSNNK 795

Qy 785 ISTLSNQSFSNMNTQLTLTLLSYNLRICPPRTFDGLSKRLSLSHGNDISVVPGEAFNDL 844
 Db 796 ISSLSNSSFNMNSQLTLTLLSYNALQCTPPPLAFQGLSKRLSLSHGNDISTLQECIPADV 855

Qy 845 SALSHLAIGANPLYCDONMQLSDNWKSEYKEPGIARCAAGPEMADKILLTTPSKFTQ 904
 Db 856 TSLSHLAIGANPLYCDCHLRLWSNWKTYKEPGIARCAAGPEMADKILLTTPAKFTQ 915

Qy 905 GPVDWNLAKNCPCLSNPKNDGTCSNDVDFRTCTPGFKGQCDVPIHACISNPKCK 964
 Db 916 GPPSLAVQAKDCPLSSPCNQGTCHNPLEVTRCTPGSYKRNCSYSSSSPCGN 975

Qy 965 GTGCHKEEGEDGFWICADGFECEVNDCCNDENNSTCVGGINNYTCLCPPEY 1024
 Db 976 GTGCHKEEGEDGFTSCSPFSGELTGMNTDCCVHDCVNGVCGVGINNYTQCPQLQ 1035

Qy 1025 TELCEELDPAQDANLPCQDHSKCLIPKPGKTCPTGYVEHGDIDFDQDCKNKGK 1084
 Db 1036 TGRACQLVDVPCFDLANCPQHAQCVGTPEGFCBVPSTGDNCKMGGCKDCKBQCG 1095

Qy 1085 ABCTDANWNTICPCBEGYSLCFEFSPPMVLPRTPSCDNFQDQNGAQCIVRIPEIQC 1144
 Db 1096 MQCVDINSTALCAGTSGQLCEIPP---APANS-CESTEQNGANCVQDGSFVQCL 1151

Qy 1145 PGTGCECKLKSVMFNKESLQIPSAKVPQTNTLIQATDESSILLHKGNDHIAV 1204
 Db 1152 PGTGCECKLKSVMFNVDROTILPTDQLWPRANTLYTGYAEENILLHKGNDHIAV 1211

Qy 1205 ELHGRVRASTDGSHPASAITSVETINDGNFHVIELLADQSLSDVGGNPKIITNLS 1264
 Db 1212 ELHGRVRASTDGSHPASAITSAETINDGNFHVIELVQDMVNLIDGSGPNTMONG 1271

Qy 1265 KQSTLNFDSPLVGMGKSNVASLRQAQNGTSPHGCIRNLIINSELQDFKVPWMTG 1324
 Db 1272 KHTLNSAEPFLVGMGPDVNSAARLQNLGNTSPHGCIRNLIINSELQDFKTPWMPG 1331

Qy 1325 ILPGCEPCCHKVCAHGTQCPSSQAFTCEQCEGWMGLCDQRTNDCPLGNCKVHGTCLPI 1384
 Db 1332 VVPGCEPCCKLVLGAIQCPNATPGVPCRCAGWGLGLCDQPDVPGCHGCKVHGKVPCL 1391

Qy 1385 NAFSTSKCLBGHGVGKLCDEEDLNPQAIKHKGRKSLGLQPYCBSSGTYGSDC 1444
 Db 1392 DALATSQDQGTGSLCNGVAVARECGGLGHLGHGQASATGABGVCSFGSGELCE 1451

Qy 1445 REISRCGERIRIDYKQOQYACQTTTKVSRLECRGCAAGCGPLSKRKYSFECTD 1504
 Db 1452 QESRCGERDPVDFHVRQGYALCQTRPLSWECRCAGCPGCGGCLLKRRLKLPFCS 1511

Qy 1505 GSSFYDEVKVKKSCTRC 1523
 Db 1512 GTSFAEYKFPKCGCAPC 1530

RESULT 3
 742626

secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)

N:Alternate names: neurogenic extracellular slit protein

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000

C:Accession: 742626

R:Holmes, G.P.; Negus, R.; Burrige, L.; Raman, S.; Alger, E.; Yamada, T.; Little, M.H. Mech. Dev. 79, 57-72, 1998

A:Title: Distinct but overlapping expression patterns of two vertebrate slit homologs in

A:Reference number: 222177; MID:9927238

A:Accession: 742626

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1025 <HOL>

A:Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PID:AA04345.1

C:Genetics:

A:Gene: Slit2

Query Match 66.5%; Score 5530; DB 2; Length 1025;
 Best Local Similarity 95.9%; Pred. No. 0;
 Matches 983; Conservative 25; Mismatches 17; Indels 0; Gaps 0;

Qy 501 ACPCRCRCBGTVDCSNQKLNKIPHTPQVTAELNNNEPTVLEATGIFKLPQLAKIN 560

Db 1 ACPCRCRCBGTVDCSNQKLNKIPHTPQVTAELNNNEPTVLEATGIFKLPQLAKIN 60

Qy 561 FSNKNTIDEGAFEGAGVGNWILLTNRLENVQHKMPKGLSELTMLKLSNRWISCTVGD 620

Db 61 FSNKNTIDEGAFEGAGVGNWILLTNRLENVQHKMPKGLSELTMLKLSNRWISCTVGD 120

Qy 621 SFGLSSVRLSLVDNQITTVAGAFDPLASLTMLANPFWNCNLAHGLWGLKRRKT 680

Db 121 SFGLSSVRLSLVDNQITTVAGAFDPLASLTMLANPFWNCNLAHGLWGLKRRKT 180

Qy 681 VTGNPCRCQPTFLKEIPIDQVAIQDPTCDGNDNNSCSPLSRCPETCLDTVVRCSNG 740

Db 181 VTGNPCRCQPTFLKEIPIDQVAIQDPTCDGNDNNSCSPLSRCPETCLDTVVRCSNG 240

Qy 741 LKVLKPGIPROVTELYLDGNQFTVLPKLSNKHLLTDLSSNRISTLSNGSFSNHTQL 800

Db 241 LKVLKPGIPROVTELYLDGNQFTVLPKLSNKHLLTDLSSNRISTLSNGSFSNHTQL 300

Qy 801 TLLSYNLRICPPRTFDGLSKRLSLSHGNDISVVPGEAFNDLSALSHLAIGANPLYCD 860

Db 301 TLLSYNLRICPPRTFDGLSKRLSLSHGNDISVVPGEAFNDLSALSHLAIGANPLYCD 360

Qy 861 CNMQWLSNWKSEYKEPGIARCAAGPEMADKILLTTPSKFTQGPVDWNLAKNCPCL 920

Db 361 CNMQWLSNWKSEYKEPGIARCAAGPEMADKILLTTPSKFTQGPMDITIQAKNCPCL 420

Qy 921 NPKNDGTCSNDVDFRTCTPGFKGQCDVPIHACISNPKKHGTCKLBGEDEGFC 980

Db 421 NPKNDGTCSNDVDFRTCTPGFKGQCDVPIHACISNPKKHGTCKLBGEDEGFC 480

Qy 981 ICADGFECEVNDCCNDENNSTCVGGINNYTCLCPPEYTGELCEELDPAQDQL 1040

Db 481 ICADGFECEVNDCCNDENNSTCVGGINNYTCLCPPEYTGELCEELDPAQDQL 540

Qy 1041 NPQCBKSLCITLPKFGKCDCTPGYVGEHCDIDFDQDCKNKNAGCTDAVNGTCTCPE 1100

Db 541 NPQCBKSLCITLPKFGKCDCTPGYVGEHCDIDFDQDCKNKNAGCTDAVNGTCTCPE 600

Qy 1101 GYSLGFCFSPMVLPRTPSCDNFQDQNGAQCIVRIPEIQCPLGTQCECKLKSVMFN 1160

Db 601 GYSLGFCFSPMVLPRTPSCDNFQDQNGAQCIVRIPEIQCPLGTQCECKLKSVMFN 660

Qy 1161 INKESYLAIPSAKVPQTNTLIQATDESSILLYKGDHIAVELYGRVRASTDGS 1220

Db 661 INKESYLAIPSAKVPQTNTLIQATDESSILLYKGDHIAVELYGRVRASTDGS 720

Qy 1221 PASAIYSYETINDGNFHVIELLADQSLSDVGGNPKIITNLSQSTLNFDSPLVGMG 1280

Db 721 PASAIYSYETINDGNFHVIELLADQSLSDVGGNPKIITNLSQSTLNFDSPLVGMG 780

Qy 1281 PGKSNVASLRQAQNGTSPHGCIRNLIINSELQDFKVPWMTGILPGCEPCCHKVCAH 1340

Db 781 PGKSNVASLRQAQNGTSPHGCIRNLIINSELQDFKVPWMTGILPGCEPCCHKVCAH 840

Qy 1341 TCQPSQAQFTCEQCEGWMGLCDQRTNDCPLGNCKVHGTCLIPAFSTSKCLBGHGV 1400

Db 841 TCQPSQAQFTCEQCEGWMGLCDQRTNDCPLGNCKVHGTCLIPAFSTSKCLBGHGV 900

Qy 1401 LCOEEDLNPQAIKHKGRKSLGLQPYCBSSGTYGSDCSREISRCGERIRIDYK 1460

Db 901 LCCDEEDLNFQINQKNGKRLSGVQPYCECNSGFTGDCDREISCGRIRIYTK 960

Qy 1461 QQCAATACQTKRKVSLRCGGCAGGCGCPSLRKRRKTSFECTDGSFVDEKVKVCG 1520

Db 961 QQCAATACQTKRKVSLRCGGCAGGCGCPSLRKRRKTSFECTDGSFVDEKVKVCG 1020

Qy 1521 TRCVS 1525

Db 1021 ARCAS 1025

RESULT 4

A36665

slit protein 1 precursor - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 30-Apr-1991 sequence_revision 30-Apr-1991 #text_change 21-Jul-2000

C:Accession: A36665; S13523

R:Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.

Genes Dev. 4, 2169-2187, 1990

A:Title: slit: an extracellular protein necessary for development of midline glia and cc

Reference number: A36665; MUID: 91099665

Accession: A36665

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1480 <RD>

A:Cross-references: GB:X53959; MID:g8614; PID:CAA37910.1; PID:g6615

C:Genetics:

A:Gene: FlyBase:slit

A:Cross-references: FlyBase:FBgn0003425

C:Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glyc

C:Keywords: alternative splicing

F:66-91/Domain: proteoglycan amino-terminal homology <PAR1>

F:101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>

F:125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>

F:149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>

F:173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>

F:197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>

F:228-272/Domain: proteoglycan carboxyl-terminal homology <PCS1>

F:288-313/Domain: proteoglycan amino-terminal homology <PAR2>

F:323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>

F:347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>

F:371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR8>

F:395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR9>

F:419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>

F:450-494/Domain: proteoglycan carboxyl-terminal homology <PCS2>

F:512-537/Domain: proteoglycan amino-terminal homology <PAR3>

F:547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>

F:572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>

F:596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>

F:620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>

F:651-695/Domain: proteoglycan carboxyl-terminal homology <PCS3>

F:708-733/Domain: proteoglycan amino-terminal homology <PAR4>

F:743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>

F:767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>

F:791-814/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>

F:815-838/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>

F:846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>

F:1028-1061/Domain: EGF homology <EGF>

F:1068-1099/Domain: EGF homology <EGF>

F:1115-1148/Domain: EGF homology <EGF>

Query Match 41.9%; Score 3486; DB 2; Length 1480;
Best Local Similarity 43.8%; Pred. No. 2.9e-196;
Matches 660; Conservative 275; Mismatches 45; Indels 116; Gaps 22;

Qy 18 CPAQCSGSDVDCGHLARSLVPRNIPRTERLDGNNNITRIITKDPAGLRHVLQML 87

Db 73 CPNVCSTGLVDCSHRLGTVPRKISADVERLELQGNLVIVITETPQLRLKMLGT 132

Qy 88 ENKISTIERGAFQDLKELRLNLRNHLQFLPELFGTAKLYALDLSENQIAQPKAF 147

Db 1156 FVHNNSFVELPLRTPKANVIVTFSSAGNGLIMYDQDAHLAVELFNGRIYVSDVGN 1215

Db 133 DNQIHTIERNFQDLVSL-----RDLSNNVITVGRVRF 168

Qy 148 RGAVIDNKLQLDYQISCIEDAGFRALDLVLNNNNITRLSVAFNHMFKITFLRL 207

Db 169 KGQASRLSLQNNQITCIDEHAFGLVELEITLNNNNITSLPHNPGGLGRLLRLS 228

Qy 208 SNNLVDCCHLAWLSDWLKRPRVGLYTCQMGPSHLGRHVAQKRFVCSDEEGHQSF 267

Db 229 DNPFAQDCHLSNRLSATPLAPTRCQSPSQLGQVADLHDQFKSCGITE----- 283

Qy 268 MAP-SCVSLH-CPAQCSTSNIVDCRGKGLTEIPNLPETITEIRLQNTKIVPPGAFS 325

Db 284 HAPMECAENSQPCPCRCAGDVCRESLTSVPVLPDPTDVTDLQNPITELPKPSFS 343

Qy 326 PYKLRLRLDSNNQISELAPDAGFLRSLNLSVLGNKITELPKSLFGLSLQLLALA 385

Db 344 SFRLRLRLDSNNNISRIADHLSGLKQTLTVLYGNKIDLPSPGVGGLSLRLLLLA 403

Qy 386 NKNLNLRLVADQDLNMLNLSYDNKIQITAGTFTSPRLAQTHMLAQPFICDCHLWL 445

Db 404 NELSCIKRDPDLHLSLSLSLSDYNTGISLANGTPAMKSMVTHLAKMPTICDNLWL 463

Qy 446 ADYLPNPTFTSGARCTSPRLNKRIGQIKSKKFCSTEDYSKSLSCDPAFLACPE 505

Db 464 ADYLPNPTFTSGARCTSPRLNKRIRSLREKFKCSWGE-LAMKLSGPCDCDPCAP 522

Qy 506 CRCEGTTVDCSNQKLNIPEHIFQTTAEALNNEFVLEATGFKPKLPQAKINFSNN 565

Db 523 CBCEGTTVDCSNQKLNIPEHIFQTTAEALNNEFVLEATGFKPKLPQAKINFSNN 582

Qy 566 ITDIEGAFEGAGVNEILLTSNRLNVLQVHNGKLSLKLMLNRLNRTVCGNDSFGL 625

Db 583 LTGIEPNAFEGAGVNEILLTSNRLNVLQVHNGKLSLKLMLNRLNRTVCGNDSFGL 623

Qy 626 SSVRLSLSDYNTGISLANGTPAMKSMVTHLAKMPTICDNLWL 685

Db 624 ----LNLDTYQICVMPGSEHLNLSLNLASLPFNPCNHLANFACVYKSLGANGAA 678

Qy 686 RQKPYFLPEIKQIDVAIQDFTCDGNDNDCSPLRCEBTCTDITVTVCSNKLKVL 745

Db 679 RQKPYFLPEIKQIDVAIQDFTCDGNDNDCSPLRCEBTCTDITVTVCSNKLKVL 737

Qy 746 KGIPRDTVELTDLGNQFTLVPR-LSNYKRLTLIDLSNRLTSSQFSNQLTLTL 804

Db 738 KGIPRDTVELTDLGNQFTLVPR-LSNYKRLTLIDLSNRLTSSQFSNQLTLTL 797

Qy 805 SYNLRLCPPIPTFGKLSLALLSGNDISVWPEGAFNLSLALSHLAINPLVCDNM 864

Db 798 SYNLRLCPPIPTFGKLSLALLSGNDISVWPEGAFNLSLALSHLAINPLVCDNM 857

Qy 865 WLSDWYSKTEPIGARCAGPGEMAKLLITPSKPTCQGPVYVNLKMLNCLSPNPK 924

Db 858 WFLSDWYSKTEPIGARCAGPGEMAKLLITPSKPTCQGPVYVNLKMLNCLSPNPK 917

Qy 925 NDGTCNSDVPDVTCTCPYFGKDCQDPVHACISNPKHGKTHKEGEEDFCWICAD 984

Db 918 NDAQVALPQREYQCLCPQYHGKHCERMDIACQGNPCNNATCVL-EEGFSQDQAP 975

Qy 985 GFEGECENNVNDC-EDNCCENNSTVCGNINVTCLCPPEYTGELCEKDLQCAQNL 1043

Db 976 GTGARGENTDLDGELKQNNACICDGVSTKCEQPGSGECFTDQICSPFEPNK 1035

Qy 1044 QHDSKILTPGFKGCDTCTPGYVGHCDIDPDQDNKNCNAGCTDAMVGTICPCBYS 1103

Db 1036 ANKAGCMQHTFYSCDQAGFGNCTNDIDQNMCGNGTCDVGINDTQCPDPTD 1095

Qy 1104 GLPCEFP--FWPLPSTPCSDPQCNQA--QCIWRINPEICQLPGYGEKCEKLSVW 1159

Db 1096 GXCEGHMIMNMPQTPSPQCNHCKHGVCPQNAQSDYLRCCHPTGKCEKLSVW 1155

Qy 1160 FINKSEYLIQPSAVRPQTNIQTATDESGILLYKDKDHIAVELYAGRVAVDTGS 1219

Db 1156 FVHNNSFVELPLRTPKANVIVTFSSAGNGLIMYDQDAHLAVELFNGRIYVSDVGN 1215

Qy 1220 HPAISVTSVETINDGNFHVILLADQSLSLVDGNGPKITNLKSTQSLNDFSLPVGG 1279
Db 1216 HPSVTHSPFVADGKHAVELLAKKNTLRVGRASLINSNGSNVLYLKTPTMPLGG 1275

Qy 1280 MPAKSNVSLRGAQPGQNTSFHGICRLNLYNSELDQKQVPMQGLPGCECHKKVCA 1339
Db 1276 LPVDPQAQYAKNWRILNLTSPKCGKGVYINHLKLVDFGNAQQKITPGC-----ALLE 1329

Qy 1340 GTQCPQSAGRTGCEQBGWNG-PLCDQRTNDPLCNKCVHGT-CLP-IMA-FSTSKCL 1394
Db 1330 GQEQQEE-----DDQDFMDETHPIKEPVDPLENCKRGSGCVFNSNADQYQCKCK 1383

Qy 1395 BGGHGVLCDEEDLFPQCAIKCKHKGSLGQPYPCSCSYTGDSCDEISCRGERI 1454
Db 1384 HQGRGYTCQDGBSTPE-----PTVTAAS-----TCRKEQV 1414

1455 RHYTKQGGTAACQTTTKVSRLECRGGCAGCGCGPLSRKRRKYSFCTDGGSPFVDEYK 1514
1415 RYTYEND-----CRSROPLKAYCGGC-GNCCAAKIVRRKRVMSNNRKIKNDI 1469

Qy 1515 VVKGGCTR 1522

Db 1470 VVKGGCTK 1477

RESULT 5

B36665

slit protein 2 precursor - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 19-May-2000

C:Accession: B36665

R:Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.

Genes Dev. 4, 2169-2187, 1990

A:Title: slit, an extracellular protein necessary for development of midline glia and cc

A:Reference number: A36665; MUID: I0199665

A:Accession: B36665

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1469 <not>

A:Cross-references: GB:X53959

C:Genetics:

A:Gene: FlyBase:slit

A:Cross-references: FlyBase:FBgn0003425

C:Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glyco

366-91/Domain: proteoglycan amino-terminal homology <PAR>

F:101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR1>

F:125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR2>

F:149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR3>

F:173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR4>

F:197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR5>

F:228-272/Domain: proteoglycan carboxyl-terminal homology <PCSL>

F:288-313/Domain: proteoglycan amino-terminal homology <PAR>

F:323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR6>

F:347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR7>

F:371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR8>

F:395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR9>

F:419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR10>

F:450-494/Domain: proteoglycan carboxyl-terminal homology <PCSL>

F:512-537/Domain: proteoglycan amino-terminal homology <PAR>

F:547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR11>

F:572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR12>

F:596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR13>

F:620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR14>

F:651-695/Domain: proteoglycan carboxyl-terminal homology <PCSL>

F:708-733/Domain: proteoglycan amino-terminal homology <PAR>

F:743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR15>

F:767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LAR16>

F:846-890/Domain: proteoglycan carboxyl-terminal homology <PCSL>

F:1028-1061/Domain: EGF homology <EGF>

F:1068-1099/Domain: EGF homology <EGF2>

F:1115-1148/Domain: EGF homology <EGF1>

Query Match 41.8%; Score 3475.5; Db 2; Length 1469;
Best Local Similarity 43.5%; Pred. No. 1.2e-195;
Matches 656; Conservative 274; Mismatches 451; Indels 127; Gaps 21;

Qy 28 CPAQCSGSGTVCHGSLARSPVPRNIPRTERLDLNGNNTIRITKTFDAGLAHLRLVQLM 87
Db 73 CPVVCSTGLVCHGSLARSPVPRNIPRTERLDLNGNNTIRITKTFDAGLAHLRLVQLM 132

Qy 88 ENKISTIEBGAQDLKELERLNRNHLQFPELLFGLTAKLYELDLSENQIAPKPAF 147
Db 133 DNQIHTERNSPDLVSL-----RLDISNNVITTVGRKVF 168

Qy 148 RGAVIDNQLQLDNLQISCTEDGAFRALDLVLLNNNNITRLSVAFNMHMLPFLRL 207
Db 169 KGAQSLRLQLDNLQISCTEDGAFRALDLVLLNNNNITRLSVAFNMHMLPFLRL 228

Qy 208 SNLKYCDCHLAWLSDLRKPRVGLYTCQMGSLHGRNVAQVKEKVFSCDEEGHQSF 267
Db 229 DNFPADCHLAWLSDLRKPRVGLYTCQMGSLHGRNVAQVKEKVFSCDEEGHQSF 283

Qy 268 MAP-SSVYL-CPAQCSTNNVDCRGKGLIETPNLPTTELRLQNTKIVIPGAFS 325
Db 284 HAMDEGAGNSCFPCRCADGVCCREKLSVFPVPLDPTDVLRLQVITELPFSFS 343

Qy 326 PYKLRLRLDNLNNISLAPDAGQLRSLNLVLYGNKTEPLSLPGLFLQLLLNLA 385
Db 344 SPRLRLRLDNLNNISLAPDAGQLRSLNLVLYGNKTEPLSLPGLFLQLLLNLA 403

Qy 386 NKLNLVLDAPQLNHLNLLSLDNLKLTAKGTSPRLAQTTHLQANPFCIDCHLML 445
Db 404 NEISCKEDAPQLNHLNLLSLDNLKLTAKGTSPRLAQTTHLQANPFCIDCHLML 463

Qy 446 ADTLNPNISGACSTSPRLANRILQISKKFRCSTGSDTSLSGDCPADLACEK 505
Db 464 ADTLNPNISGACSTSPRLANRILQISKKFRCSTGSDTSLSGDCPADLACEK 522

Qy 506 CRGGSTVDCSNQMLKIPHIPTAELRANNEFVLEATGFKPLQALIKVSNK 565
Db 523 CRGGSTVDCSNQMLKIPHIPTAELRANNEFVLEATGFKPLQALIKVSNK 582

Qy 566 ITDIEBGAFAASGWEILLTNRLNQLQWQWFGSLSLKTLMLNRLNRTVGNDSFGL 625
Db 583 ITDIEBGAFAASGWEILLTNRLNQLQWQWFGSLSLKTLMLNRLNRTVGNDSFGL 623

Qy 626 SVRLLSLYDMLITVAPADQTLRLSLNLNLPNCNLYLAWLGNKRRIVTGNP 685
Db 624 ----LNDMLQISLWVPSFHLNLSLNLNLPNCNLYLAWLGNKRRIVTGNP 678

Qy 686 RQKPYFLKLEPIQDVAIQDTCDDGNDNCSPLRCPTCTCTDVTVCSSNGLKVL 745
Db 679 RQKPYFLKLEPIQDVAIQDTCDDGNDNCSPLRCPTCTCTDVTVCSSNGLKVL 737

Qy 746 NGIPROVTELYDQNTFLVPE-LSNYKHLTLDLNLSNLTSLNSQFSNQLTLTL 804
Db 738 NGIPROVTELYDQNTFLVPE-LSNYKHLTLDLNLSNLTSLNSQFSNQLTLTL 797

Qy 805 STNKLCPPTFPDGLSLRLSLHGNISVVPAGFANLSALSHAGNPLCYDMMQ 864
Db 798 STNKLCPPTFPDGLSLRLSLHGNISVVPAGFANLSALSHAGNPLCYDMMQ 857

Qy 865 WLSWDNYSYKSGPIAGACGPMADKLLITPSKFTQCGPVDNLAKNPLSNPK 924
Db 858 WLSWDNYSYKSGPIAGACGPMADKLLITPSKFTQCGPVDNLAKNPLSNPK 917

Qy 925 NDCGTCSNPDVPTCTCPYGGQDCCVTHACISNPKHSGTCHLKGDEEDGFWICAD 984
Db 918 NDCGTCSNPDVPTCTCPYGGQDCCVTHACISNPKHSGTCHLKGDEEDGFWICAD 975

Qy 985 GPGHCEENFVND-ENDCENNSTVGVNINLTCLPPEYTGELCEKDLCDALNFC 1043
Db 976 GPGHCEENFVND-ENDCENNSTVGVNINLTCLPPEYTGELCEKDLCDALNFC 1035

us-09-540-245a-2.rpr

Dh	192	DCNQNLKVRIPSHLSPVTDLRANDNEVSVLEATGPKFLPNKRLKLNKNKIKVREGA	251
Qy	574	PEAGSGWELLTSLNLEPNQWKMFKLESLKTLALRSNRTICVGNDSIKGVSVLLSL	633
Dh	252	FGAASVQELMLTGQLETVHGKPVFLSGKLTALRSNLIGCVSNDTFAGLSVLLSL	311
Qy	634	YDNQITTVAPGAFDLHSLSLT	655
Dh	312	YDNRITITTPAGFTILSVLSLT	333

epidermal growth factor-like protein slit - fruit fly (*Drosophila melanogaster*) (fray)
 C:Species: *Drosophila melanogaster*
 C:Date: 28-Feb-1990 fsequence_revision 28-Feb-1990 ftext_change 11-Jan-2000
 C:Accession: A31640
 R:Rothberg, J.M.; Hartley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
 Cell 55, 1047-1059, 1988
 A>Title: slit: An EGF-homologous locus of *D. melanogaster* involved in the development
 A:Reference number: A31640; MUID:89077533
 A:Accession: A31640
 A:Molecule type: DNA
 A:Residues: 1-530 <RDF>
 A:Cross-references: GB:M23543; NID:g340939; PID:g514357
 C:Genetics:
 A:Gene: FlyBase:slit
 A:Cross-references: FlyBase:FBgn0003425
 A:Introns: 470/3
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: growth factor
 F:148-181/Domain: EGF homology <EGF>
 F:188-219/Domain: EGF homology <EGF2>
 F:215-268/Domain: EGF homology <EGF3>

Query Match 13.4%; Score 1115.5; DB 2; Length 530;
Best Local Similarity 39.0%; Pred. No. 5.5e-58;
Matches 208; Conservative 102; Mismatches 199; Indels 25; Gaps 11;

QY 1303 CIRNLTNGELQDPQRVPMQTGILPGCEPCHRRVCAGHTCQPSQAGFTCECQEGWMG-- 1361

DB 418 CMKLVWINKLVDPGNAQRQQRITPGC-----ALLEGEQEEEE-----DDEQDFMDET 465

Qy 1503 TDGSSFVDEVEKVVKCGC 1520
:|::: |: :||
Db 574 KNGTTRKISTVHLIRCC 591

RESULT 9
 A24420
 notch protein - fruit fly (*Drosophila melanogaster*)
 N:Alternate names: neurogenic repetitive locus protein
 C:Species: *Drosophila melanogaster*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A24420; A24768; S09358; A05267
 R:Kidd, S.; Kelley, W.R.; Young, M.W.
 Mol. Cell. Biol. 6, 3094-3108, 1986
 A:Reference number: A24420; MWID:87056424
 A:Accession: A24420
 A:Molecule type: DNA
 A:Residues: 1-2703 <KID>
 A:Cross-references: GR:Q03508; MWID:g157991; PIDN:AAA82725.1; PID:g157993
 R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
 Cell 43, 567-581, 1985
 A:Reference number: A24768; MWID:85079539
 A:Accession: A24768
 A:Molecule type: mRNA
 A:Residues: 1-48, '1', 50-118, 'R', 120-230, '1', 232-256, 'N', 258-266, 'A', 268-872, 'R', 874-9
 A:Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2
 R:Taute, D.

Nucleic Acids Res. 11: 6463-6471, 1989

ATitle: Hypervariability of simple sequences as a general source for polymorphic VN

A:Reference number: S09358; MUID:89385974

A:Accession: S09358

A:Molecule type: DNA

A:Residues: 2505-2551, 'QQQQ', 2552-2576, 'E', 2578-2604 <TAND

R:Wharton, K.A.; Tiedobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.

Cell 40, 55-62, 1985

ATitle: opa : a novel family of transcribed repeats shared by the Notch locus and

A:Reference number: A05267; MUID:85099329

A:Accession: A05267

A:Molecule type: DNA

A:Residues: 2504-2576, 'E', 2578-2611 <KHA2>

C:Genetics:

A:Gene: notch; opa

A:Cross-references: FlyBase:FBgn0004647

A:Map position: 8.96-8.96

A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

C:Keywords: differentiation; tandem repeat; transmembrane protein

F:27-43/Domain: transmembrane \$status predicted <TM1>

F:297-328/Domain: EGF homology <EGX1>

F:530-561/Domain: EGF homology <EGF1>

F:568-599/Domain: EGF homology <EGF>

F:988-1019/Domain: EGF homology <EGX2>

F:1064-1095/Domain: EGF homology <EGF3>

F:1187-1218/Domain: EGF homology <EGX3>

F:1746-1762/Domain: transmembrane \$status predicted <TM2>

F:1950-1982/Domain: ankyrin repeat homology <AN1>

F:1983-2015/Domain: ankyrin repeat homology <AN2>

F:1988-2004/Domain: transmembrane \$status predicted <TM3>

F:2017-2049/Domain: ankyrin repeat homology <AN3>

F:2050-2082/Domain: ankyrin repeat homology <AN4>

F:2083-2115/Domain: ankyrin repeat homology <AN5>

F:2538-2568/Region: glutamine-rich

F:2538-2568/Domain: neurotensin repetitive element \$status predicted <CPA>

A;Molecule type: DNA

R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985.

A;Accession: A05267

A;Residues: 2504-2576,'E',2578-2611 <WHA2>
C:Genetics:

A;Cross-references: FlyBase:FBgn0004647
A;Map position: 8.96-9.36

C;Keywords: differentiation; tandem repeat; transmembrane protein

F;297-328/Domain: EGF homology <EGX1>
F;530-561/Domain: EGF homology <EGF1>

F;988-1019/Domain: EGF homology <EGX2>
F;1064-1095/Domain: EGF homology <EGF3>

F;1746-1762/Domain: transmembrane #status predicted <TMM2>
F;1950-1982/Domain: ankyrin repeat homology <AN1>

F;1988-2004/Domain: transmembrane #status predicted <TM3>
F;2017-2049/Domain: ankyrin repeat homology <AN3>

F;2083-2115/Domain: ankyrin repeat homology <AN5>
F;2538-2568/Region: glutamine-rich

(continued)

Matches 262; Conservative 93; Mismatches 336; Indels 202; Gaps

Qy 707 TCDDGNDNDS¹CSPLSRCPTECT---CLDTVVRCSNKG²LKVLPGKIPRDVTELYLDGNQFT 763
||| ||| : ||| | ||| : : | | |

DD FORM 10-67

397 1C1D813B11Y KOFFARLONTCQUDTDCXHQHDFE TCGRDNHCJNIB 399

Qy 764 LVPKSLNYSKHLTL-----IDLSNNRISTLSNQSFNNMQLLLLSYNLRICPIPTFD 818
Db 354 -----GSTSCICVWAGLDCSNWTDCKQAQCFYAT-----CI-----D 389

Qy 819 GLASLLSLSLGNDISVVPBGAFNLSALSHL--ATGANPLYDC--CNMQLSD-----WV 870
Db 390 GVGSTFQCTKXK-----TGLCHLLDACSINPCRADACICTDTPINGSIACSC 437

Qy 871 KSEYK-----EPGIARC--AGPGEMADKILLTTSKFKTCQ---GV--VDVNLAKNCPL 919
Db 438 ATGKGVDCSEIDECDOGSQCEHNGICVNTPGSVKNCISQGTGPRCETNI-----NDC 493

Qy 920 SNPKCKMGTSDGSPVDFYRCTPYGFKGQDCDPIRACINPCKHGGTCKLAGEBGF 979
Db 494 SHPCQBGSCLDGPTG-RVCNMPGFTGTQCEIDIDCQSNPLCNDGTCKD-----INGFK 549

Qy 980 CIGADGPEGENEYVNDCCDNDENNSTCVDSINNYTCLPPFTYGLCEKELDPAQD 1039
Db 550 CSCALGPTGARQINIDCQSPQRNRRIGCHDSIAGSYCECPGYTGTSCHEINDC--D 607

Qy 1040 LNPCKMGSCKILTPGFKCQDCTPGYVGEBCIDFDDQDNKCKNAGACTAIVNGYTICP 1099
Db 608 SNPC-HRGKCIDVNSFKCLDQPGYTGICQKINDECSNPQDQBGQDQWGSYTCQ 666

Qy 1100 BEYSLGFCFSPPLWLPRTSPCDNFDQGAQVCIENPQICLPQCEKELKSVN 1159
Db 667 AGTSGKNCVW-----VNECHSNPCNMGATCIDGINSYKQVQVPGTQHXK----- 714

Qy 1160 FIMKESYLIQISAKVRPQTNITLIQIATDEDSGILLYKGDHIAVELYGRVRASTDS 1219
Db 715 --WVDECISSPCA-----KNGVICDQWG--TK-----CECPG-----FDD--A 748

Qy 1220 HPASAI-----YSVETINDGNPHIVELLADQSLSLSDVQGNKPII--TWLSRQSTLMD 1272
Db 749 HCLSDVDECAASNPCVNEGRCE-----DGNFICHCPPGTGRCALD 793

Qy 1273 -----SPLVWG-HPGKSNVASLRAPQNG-----TSFHCINLITL-----SELDQ 1317
Db 794 BCSSNPQGGGTCTDKLAFSCQCMPTGTGCKEYINDDCVNPGNGGTICDKWNGTK 853

Qy 1318 --KVPMTGILPGCE-----PCHKVCAR-GTCQPSQ--AGPCRCQBGWGLCDQRTND 1369
Db 854 VKKPTG--KCESKMDPCASNCKNRAKCTPSNLFDFSTCKLGTIGYCDDEIDE 910

Qy 1370 PCIGNLCVWG-TCLPIAFSTCKCLEBGGVLCDDEEDLPNQCINCKHGRKSLG 1428
Db 911 CSLSSPCRNAGSLNVPG-STYCLCTKYGYGRDCAINTD-----DCASPQNGGTCLD 966

Qy 1429 QPICESSSGTGDSCREIS-----CR-GERIRDYTQKQ-----GTAA--CQTKKYS 1474
Db 967 DYSCLSDVDFGQCEETDINECLSPQCGNGATCQSYVNSTCTCPLGFSINQYND 1024

Qy 1475 RLBC-RGGCA-GGCGPLSKRRKYSPECTDGSFVDEVEKYKCGTCRVS 1525
Db 1025 --DCTESSCLNGSCIDING-----YNGSCLAGSNGACQYKLNKDNFCLM 1071

RESULT 10
A56136

jagged protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 11-Jan-2000
C:Accession: A56136
R:Lindsell, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.
Cell 80, 909-917, 1995
A:Title: Jagged: a mammalian ligand that activates Notch.
A:Reference number: A56136; MUID:95211842
A:Accession: A56136
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1220 <LIN>
A:Cross-references: GB:L38483
C:Superfamily: unassigned EGF-related proteins; EGF homology

F:379-410/Domain: EGF homology <EGF1>
F:492-523/Domain: EGF homology <EGF>
F:634-665/Domain: EGF homology <EGF2>

Query Match 9.2%; Score 768; DB 2; Length 1220;
Best Local Similarity 25.4%; Pred. No. 3.4e-37;
Matches 197; Conservative 101; Mismatches 246; Indels 232; Gaps 32;

Qy 853 GANPLYDC-----CNMQL-----SDWKSEYKEPGIARCGAGEMADKILLTT 896
Db 257 GWQGLCKDCIKPHPGCVHGTCKNEPQCLCEINW-----GGQLCKR----- 296

Qy 897 PSKFTQCGPVDVNLAKNPLSNPKNDGTCNSDPVDFYRCTPYGFKGQDCDPIHA 956
Db 297 -----DLAVCTGHQPL-----NRGTCSTGPKYQSCSPGYSGPNCIAEHA 340

Qy 957 CISNPKHGGTCHLAGEEDGFWCIDAGPBGENCEVNVDDCCDNDENNSTCVDSINNY 1016
Db 341 CLSDPCKNRGSC--KE-TSSQPEBCSPGWGTCTSTNDSCSPNCSHGGTQDQVWNG 397

Qy 1017 TCLCPFTYGLCEKELDPAQ-----DLN----- 1041
Db 398 KVCPPWQCTQDLNDECAKPCVNASCKMLASTYCDCLGHWNGQCDINCLG 457

Qy 1042 PQGDESKCLILPFGKCDCTPGYVGEBCIDFDDQDNKCKNAGACTAIVNGYTICP 1101
Db 458 QCCNDASCRDLVNYTICPPTGADGHCERDIDCASNPLAGGHCQNEINPQCLCPG 517

Qy 1102 YSLGFCFSPPLWLPRTSPCDNFDQGAQVCIENPQICLPQCEKELKSVNFI 1161
Db 518 FSGNQLQD-----DICEPNNQCAQYNRASDTPCKPCEDEGKMSHL----- 564

Qy 1162 NKESTYLQPSAKVRPQTNITLIQIATDEDSGILLYKGDHIAVELYGRVRASTDTG-- 1218
Db 565 --KDHCTTCPEIDDSCTVAM-ASMDTPGV-----RYISSNCGPHCKCSSEGGK 614

Qy 1219 -----SPASAIYSVETINDGNPHIVELLADQSLSLSDVQGNP-----KIITNLSRQSTL 1269
Db 615 TCDNCGFTGTTCHEINDCE-----GNPCTNGTCTIDGVSYKCI 655

Qy 1270 NFDSPLVWGMPG--KSNVASLRAPQNGTSPHOCIRNLIN----- 1310
Db 656 CSD-----GWEAGACENWINDCSNPCHYGGTCDRLVDYCDCKMNGWGTCKSRSDQ 709

Qy 1311 -----SELDQ-----KVPMTGILPGCECHKVKCAHGTQCP 1345
Db 710 CDEATCINNGTGYDEVDTFKCMCPGWEGTTCIARRNSCL--NPCN-----GTCVYN 763

Qy 1346 SQAGFTCEBQDGNWGLCDQRTNDPLCKNVCY-GTCLPIAFSYSCKLEBGGVLCD 1404
Db 764 GDS-FTVCCKEHWEGPCTQNTND-CSPHNYSGTCDVGGDW-YRCECAPGAPGDCRI 820

Qy 1405 EEDLPNQCINCKHGRKSLGQPYCESSSGTGDSCREISCR-----GERIRDYTQ 1459
Db 821 N---INDEQSPCAFGATCVDEINQYCIQPGHSGAKC-HEVSGRSCTMGRVILQAG 876

Qy 1460 KQGYAAQTTTKVSRLEKRGCCAGG--C-----CPLSKRRKYSPECTDGSFV 1509
Db 877 WDDCCTQ-----CLNGRVACKSWKQCPRPCKLHKGBCYNGQSCI 919

RESULT 11
A35844

Xotch protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35844
R:Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A:Title: Xotch, the Xenopus homolog of Drosophila notch.
A:Reference number: A35844; MUID:90385285
A:Accession: A35844

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-2524 <QOF>
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; BGF homology
C:Keywords: transmembrane protein
F:146-177/Domain: BGF homology <BGX1>
F:184-215/Domain: BGF homology <BGF1>
F:222-254/Domain: BGF homology <BGF>
F:456-487/Domain: BGF homology <BGX2>
F:757-788/Domain: BGF homology <BGF3>
F:1025-1056/Domain: BGF homology <BGX3>
F:1924-1956/Domain: ankyrin repeat homology <AN1>
F:1957-1989/Domain: ankyrin repeat homology <AN2>
F:1991-2023/Domain: ankyrin repeat homology <AN3>
F:2024-2056/Domain: ankyrin repeat homology <AN4>
F:2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 9.1k; Score 759; DB 2; Length 2524;
Best Local Similarity 26.8k; Pred. No. 2.8e-36;
Matches 240; Conservative 100; Mismatches 352; Indels 204; Gaps 43;

QY 733 VVPCSNKGLVLPKGP--RDVTELYLDGQPTLVPELSNHYKTLIDLDSN----- 783

DB 8 VLICS--LPLVQLGRPCQTAEKCLNGGKCEMTPGG-----TGVCLGNLPGERC 57

QY 784 ---RISTLSNQSFSNMQQLLLILSYNRLRCIPPTFDGLSLALLSLHGNDSIVV---P 837

DB 58 QFNPCTIRKQCNHFGT-----CEP-----VLQNAIDFICBCP 91

QY 838 EGFANFLSLSHL--AIGANPLYDCDNQWLSNWKVS--EYKEPIARCAQGGEMADKL 893

DB 92 VG-FIDKWLIPVNNKYNP-----CRNGCTCLLSVTEIK---CRC-PQWISQSQ 141

QY 894 LTPSPKFTQPG-----PVDVNLIAK-----NPLSNPKKNGTGNDSVD 935

DB 142 QADPCASPCANGKCLPFPIQITCKPGPGHATCKQDINCSNPKSGGGCIG-FG 200

QY 936 FYTCTCPFGKQCDQVPHACISNPKHSGTCLAHGBEDGFICADGFEGENEYV 995

DB 201 SYTCTQKQNPRTGKNCDEPTVPCNPSCLNGTIC--RQTDSDYCTCLPSPGSCHEMI 258

QY 996 DCCEDDCCENNSTCVGGINNYTCLPPEYTGELCEKRLDFCMQDNLQPNDSKILTPK 1055

DB 259 DDCPSNCRNGSTCVGVTNYCCQPPDQYCTEDVDEQCLMPNACQGGTCTNVTGG 318

1056 FKDCDTPGVGEHCDIFDQCDNCKNGARCTDAVNGYITCIPGTSGLCEFSPPMW 1115

DB 319 TNCVNVNMGDECSNIDCNAACHGATCAHRSVATPCBCPHGTLGLCHLD----- 373

QY 1116 PRTPSCFNQDQGAQICVR-IN-EPICQPLGYQKCEKLSVYNFKINESTYLPQSAK 1173

DB 374 ---NACISNPNCSNCDNPNVNGKAICTCPPTGTPAFAN-----NDVDECSLGAHPCR 425

QY 1174 VRPQNTLIQTIDEDSGILLYKSDHIAVELYGRVRSYDTGSHPASALTSYETID 1233

DB 426 GGCCTNLTGSCNCPQG---YAGPRCEIDVN-----ECLSPQ-----ND 464

QY 1234 -----GNFRIVELLALDQ---SLSGLVDGKNPKIITN--LSKQSTNDSFLVGGMP 1281

DB 465 STCLDQIGFQICMCPGEGLYCETNIDECASNPLCNCKIKDINEFDCDPTGFSGNL 524

QY 1282 GKNVSNLQAQAGQNTSPHGI--RNLYINSELQPOKVMQIGI---LPGCEPCHKKV 1336

DB 525 QQDPOBCTSTPCNGAK---CLDGPNSYTCQTEGPTHGBCDIDBCTI--DPCH--- 576

QY 1337 CABGTQPSQAGTCQBEHMGPLCDQRNDCPLGNKVGTLCLPNAFSTCKCLRG 1396

DB 577 --TGTCG-DGATFTTCLRCPTGTGLCDNDINE-CLSKPCLNGQCTQDRENGYICTPKG 632

QY 1397 HGVGLCDDEEDLWPCQAIKXKNGK--RLSGQPYCECDSSTGDSQREIS----- 1448

DB 633 TTYGVTBKID---DCASNLDGKCIKIDIGI---BCTEPGTGALCNINIDCNSP 686

QY 1449 CR-GERIDYIYK-----QGGY-----AAQCTTKVYSRLCRRGCA 1483

DB 687 CRNGGTCKQDQINGFTCVCPDGHDBNCLSVNECNSNCPICAGCHDGVNRYKDCDEAGNS 746

QY 1484 GGQC-----C---GPLSRKKRYSFECTDGGSFVDEKVNKQCTHCVS 1525

DB 747 GSNCDINNNECNSNPNMNGGTCKDMTGAVICTKAGFSNPGQNTNINECSNPNCLN 802

RESULT 12

T30201

Notch homolog protein - sea squirt (Halocynthia roretzi)

C:Species: Halocynthia roretzi

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T30201

R:Horii, S.; Saitoh, T.; Matsumoto, M.; Makebe, K.W.; Nishida, H.

Dev. Genes Evol. 207: 371-380, 1997

A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the

A:Reference number: t20775

A:Accession: T30201

A>Status: preliminary; translated from CR/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2352 <HOB>

A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25572.1

C:Genetics:

A:Gene: Notch

Query Match 9.1k; Score 757; DB 2; Length 2352;

Best Local Similarity 23.1k; Pred. No. 3.4e-36;

Matches 305; Conservative 139; Mismatches 480; Indels 396; Gaps 64;

QY 438 CDCHLWAD-----VLATNPITSGARCTSPRLANKRIQKISKFKCGTGTYSK 491

DB 96 CTQCTGPTGDCSCVLKCSNPK-SNGAGC-----EELSNFKCTCTGTGTD 142

QY 492 LSGQCFADLACEKRCBGTVDSCNKLKNIPEHPIQPTAELRNNNEFTY----- 543

DB 143 TCANDVNBCTPDICQAGT---CSNND---GGTSCSVAGFGNNGENYVDCSGHSC 195

QY 544 -----LEATGIFKRLPQLRKTFNSN---KITDIE--BGAPEAGSVN 581

DB 196 QNGATCADAVSTYDCHPAEWGTQCTI-DVDECSLSNNVAKRRLQGTGFG----- 247

QY 582 EILLTSNL-----ENQVH-----KMFKESLKTMLRSNRTICVGNDSF 622

DB 248 ---TNCVYTPTRDCSENIDDCSNVACFNHACIDQAGTPECTICPNRILCHLDAC 303

QY 623 IGLSSVLLSLSDNQIT---TVAPAG---DTLHSLSTMLLAN---P 662

DB 304 ISDPCARAGATDTPNIGHKMGCDPDGWTDKDCSDIDECISGNGPCHENGQNTDTSF 363

QY 663 NCNCTLAWGWLKRRIRVIGHNPROQPKYFLAKEIPDQVADQICDGDNDNSCPLSR 722

DB 364 EICVAGTSG-----PRCE-----TLINECEP-NP 387

QY 723 CPTECTCLDTVVRCSNKLGLVLP-KGI--PRDVTIELVD--GNQPTLVPELSNHYK 773

DB 388 CRDADATCLM---GNFKVCNMPGTGLICDEDIDECNSNPANGGCTI-DEYNATYCS 443

QY 774 --LTLTLDLSNRISTLSNQSFSNMQQLLLILSYNRLRCIPPTFDGLSLALLSLG- 830

DB 444 ALGFTGDCDSQN-IDECATPCNKNATCIDKANAY-ECECARPT---GVHCE 491

QY 831 ---NDISVVP--EGAFNDLSALSLAIGANPLYDCDNQWLSNWKVSSEYKEPIARCA 885

DB 492 TNIDCVNINHGYSGRD-----GVTFPCDLLGTEGTGKQDTDEACASPCNG 542

QY 886 EGMAKLLTLPSPKFTQGPVDVH--TLAKCN--CLSNPKCNQDNDSPDQFTYCTCP 942

DB 543 GTCTDEI-----GYTCTCPGTSGSSCEINPDQVGNQPY-GTC-VQGVWDVSCST 594

Notch protein homolog - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S18188
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID: 92111383
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <WEB>
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:987-1018/Domain: EGF homology <EFL>
F:1025-1056/Domain: EGF homology <EGF>
F:1133-1264/Domain: EGF homology <EGF>
F:1517-1549/Domain: ankyrin repeat homology <AN1>
F:1950-1982/Domain: ankyrin repeat homology <AN2>
F:1984-2016/Domain: ankyrin repeat homology <AN3>
F:2017-2049/Domain: ankyrin repeat homology <AN4>
F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 9.04; Score 747.5; DB 2; Length 2531;
Best Local Similarity 23.44; Pred. No. 1.3e-35;
Matches 229; Conservative 88; Mismatches 258; Indels 403; Gaps 39;

Qy 881 RCAGGEGMADKLLLTTPSK-----FTC-----QGVPVNHILKCNPLCS 920
Db 56 RQDQSPF---CLSTPCKNAGTCYVDHGGIVDACSCTPLGSPGLCLTPLA--NACLA 108
Qy 921 NPKCNMGTCN-----SDPV-----DFPRTCTCPY 943
Db 109 NPKRCNGTCDLTLZETKCRPPWGSKGSCQQAADPCASNPCAMGQCLPFESSYICGPP 168
Qy 944 GFGQCDQVPHACISNP---CKHGHTCLKEGE----- 974
Db 169 GFGGFCRQDYNQCSNPFCLRHGQTCNBIESTYCACRATHGTGPHCLPTVPCSPSPQ 228
Qy 975 -----EDGFWCICADGFEGENCEVNDVDCENDCENNSTCVGGINNYTCLCPET 1025
Db 229 NGGTCRPTGDTHECACLPGTAGNCEBNDVDCPNCKMGKAGCDVGYNTNCRPPET 288
Qy 1026 GELCEKLD----- 1034
Db 289 GQTCEDVDEQIMPNAQNACNAGNSHGTCNVCVNGWGDSCDSDIDDCASACFQA 348
Qy 1035 -----F 1035
Db 349 TCHDRVASVFCPCPHGTGLLCHLNDACISNPNCEGSMCDTPNVPNGKAICTCPRGTTGA 408
Qy 1036 CAQDL-----NPGQDHSKCLITPKGFKCDTPGTGYEHCIDFDQDKNKCNAGHCT 1088
Db 49 CSQVDDECALGNPCREHAGKCLNTLGSFEQCLQGTGTPRCEIDWECISNPGCDATCL 468
Qy 1089 DAVNGYTCICPBGYSGLPCFSPFPVLPRTSPNDFCQAGCAQVIRNEPIQCCLPGTQ 1148
Db 469 DQTEGFQCTCMPTGEGVTCIN-----TDBCASSPLCINGKVDKINFLQCPGPGFS 521
Qy 1149 GERC-----EKVYSVNFINKESYLQPSAKVRYPNTNITLQIATEDSGILYTKGDKHIAV 1204
Db 522 GHLQQLVDDECASTPCKNAGKLDGPNW---YTVCYTBGTITGCEVIDEDCPDPCHI 577
Qy 1205 ELRYGRVASVYDTGSEPAISVYET-IND-----GNHFIVELLA----- 1243
Db 578 GLCKDGV-ATTCCLQCPGTYGHCEININCEHSQPCRHGSTQDRDNTYLCLCLATGTP 636
Qy 1244 ----LQSLSLVDGNGPKIITNLSKQSTLMD-----SPLYVGMPGKSNVASLRQA 1292
Db 637 NCEINLDDCANFPCDSG---TCLDR---IDGYECACRPGYTGSM-CWNVIDEAGS 685
Qy 1293 PGQNGTSPHGCIARNLINSELQDQKPYMQTGL-----PGC---EPCHKV 1336
Db 686 PCRMGT-----CEDGIAGTCRCPBTHDPCLSEVNECSNP 724
Qy 1337 CAHGTQCPSSQAFTCEQCEGWSMLPCDQRTNPCLGNKVC-GTCLPINAFTSYCKCLE 1395
Db 725 CIHGAECRDLGN-GYKDCAPGSGSTNCDINNNE-CESNPNVNGTCKWMTS-GTVCTCRE 781
Qy 1396 GHGVLCDDE-----EDLFNPQOAIKXH- 1419
Db 782 GFGSPWQNTNNECASNCLAQGTCTDDVAGYKCNCPLPYTGATCEVLPACSPKNS 841
Qy 1420 GKRLSLGLQGY-CECSGYTGDSCDREIS-----CRGERIRDYKQGYGAACQTKK 1472
Db 842 GVKESEDEYSFSCVCPQWQGTCEIDINECVKSPCR-----HG-ASQNTNG 889
Qy 1473 VSRLECRGCGAGCG-----GSLRLKRRKYSPECTDG--SSFVEDY-----EKV 1515
Db 890 SYRCLCQAGTYGNCESSDIDCRPNCHN-----GSGCTDGVNAFCDCLPGFQGAFCEDI 946
Qy 1516 VKC-----GCTRCV 1524
Db 947 NCCATNPDCNAGNACTCV 964

RESULT 14

A46019
 Notch-1 protein - mouse
 N:Alternate names: notch protein
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
 C:Accession: A46019; S25144
 R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid Genomics 15, 259-264, 1993
 A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of A:Reference number: A46019; MUID:93194170
 A:Accession: A46019
 A:Status: not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-2531
 A:Cross-references: GB:Z11886; GB:S47228; NID:G288502; PIN:CAA77941.1; PID:g288503
 Note: sequence extracted from NCBI backbone (NCBI:127318)
 Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.; submitted to the EMBL Data Library, April 1992
 A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest A:Reference number: S25144
 A:Accession: S25144
 A:Molecule type: mRNA
 A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2120 <PRA>
 A:Cross-references: EMBL:Z11886
 C:Genetics:
 A:Gene: notch-1
 A:Map position: 2
 A:Note: proximal region of chromosome 2
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; BGF homology
 F:106-138/Domain: BGF homology <BG1>
 F:144-175/Domain: BGF homology <BG1>
 F:222-254/Domain: BGF homology <BG2>
 F:261-282/Domain: BGF homology <BG2>
 F:339-370/Domain: BGF homology <BG3>
 F:416-449/Domain: BGF homology <BG3>
 F:456-487/Domain: BGF homology <BG4>
 F:494-525/Domain: BGF homology <BG5>
 F:532-563/Domain: BGF homology <BG5>
 F:607-638/Domain: BGF homology <BG7>
 F:682-713/Domain: BGF homology <BG8>
 F:757-788/Domain: BGF homology <BG9>
 F:795-826/Domain: BGF homology <BG10>
 F:873-904/Domain: BGF homology <BG11>
 F:911-942/Domain: BGF homology <BG12>
 F:949-980/Domain: BGF homology <BG13>
 F:987-1018/Domain: BGF homology <BG14>
 F:1025-1056/Domain: BGF homology <BG15>
 F:1063-1094/Domain: BGF homology <BG16>
 F:1149-1180/Domain: BGF homology <BG17>
 F:1187-1218/Domain: BGF homology <BG18>
 F:1233-1264/Domain: BGF homology <BG19>
 F:1352-1383/Domain: BGF homology <BG19>
 F:1391-1425/Domain: BGF homology <BGF>
 F:1917-1948/Domain: ankyrin repeat homology <AN1>
 F:1949-1981/Domain: ankyrin repeat homology <AN2>
 F:1983-2015/Domain: ankyrin repeat homology <AN3>
 F:2016-2048/Domain: ankyrin repeat homology <AN4>
 F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 9.0%; Score 745; DB 2; Length 2531;
 Best Local Similarity 24.4%; Pred. No. 1.8e-35;
 Matches 238; Conservative 103; Mismatches 309; Indels 326; Gaps 46;

Qy 662 FNCNCLAWLGEWLKRLVITGNPRQPKYFLKEIPIDQVADQPTCDGNDNSCSPLS 721
 Db 202 YRACACNATHTG-----PHCELPY-----VPCSPQCNQ-----ATCRPTG 237
 Qy 722 RCPCTTECLDVTY-----RCSNGLKVLPLGI-----PRDVTYLL-D 758
 Db 238 DTTCACALGCFAGQCNENDDCPGNKNGGACY--DGVVYTNCRACPVTGQCTED 295

Qy 759 GNQPLVLPKELSTKHLTLIDLSSNRISTLSNQSFNSMTQLLTLLSYNRLRCIPPTFD 818
 Db 296 VDBCCQLMFWACQN-----AGTCN----- 314
 Qy 819 GLKSLRLSLGENDISVVPFAGNLSALSHAIAGNLYCDCCMQLSDWVKSEYKEFG 878
 Db 315 -----THGGYN-----CVCVNWGTGDCSEINDICA 340
 Qy 879 IARCAAGPGEHMDKLLITFSKPTQCPVD-VNLAKE-NPCLSNPKNDGTCNPDPVF 936
 Db 341 SAACPGATCTEDRV-----ASFYCECPHGRTGLLKHACISNPNBGSNCDTPVNG 394
 Qy 937 TR-CIQCYPGFG-----QD-----CDVPIR 955
 Db 395 KRITCPSSTGTGACSDVDECDLGNKCHAGKCLNTLGSFECQCLQSTTGPGCEDLVN 454
 Qy 956 ACISNPKAGSTCHLARGEEDGFWICADGPEGENCEVNVDDCENDCENNSTCVGDNW 1015
 Db 455 ECISNPQNDATCLDQIGS-----PQICMWPGEYGVTEINWDCEASFCLNHWKMDKHE 511
 Qy 1016 YTCCLPPTTGLCEKLEKLFCAQDLNPGQDSKCLITPKGPKCDCTPGYVGEHCDIFPD 1075
 Db 512 PQCCPKFQGHGLQCTVDVBCAS--TPKMGAKCLDGPNTYTCSTBSTGTGRCVDIDE 569
 Qy 1076 CDNCKCKNAGACTDANVGYTCICPGYSGLCFCEFPFPMVLPRTSPCNDFDQNGACIVR 1135
 Db 570 CDPCPCHTGS-CKDGVATPCTLCQPGTGHRC-----TINECBSQPCRHGSTCDOR 621
 Qy 1136 INEPIQCLPQSTGQKCEKLVNFWINKEYSLOIAPSRVPTNITQIATDEDSGLILY 1195
 Db 622 DNSYLCCLMLGTTGPNCE-----INLDCASNPC-----DSGTCLD 657
 Qy 1196 KGKNDHIAVEL-TRGVASTDTGSHPSASITSVETINDG-----NPIVELLAL 1244
 Db 658 KIDGTACACPEPTGSH-CNWNIDECAGSPCHNGGTCEGDIAGTCRCPEGT----- 709
 Qy 1245 DQSLSLVSDGNGKIITLNSKSTLNFDSPLYTPMGPKSNVASLRQAQPGNQTSHFICI 1304
 Db 710 -----DP-----TCLSEYNDEN-SNFCINGACRDLGLNYKCDCAFGNST----- 748
 Qy 1305 RNLYINSELQFQKRVPMQZGLPGCECHKVCYAH-GTQCPSSQAGTCEQGHGMLPLC 1363
 Db 749 -NCDINN-----NCESNPCVNGTCKDMT-SGTVCTCRBFGSPMC 788
 Qy 1364 DQRTNPGCLGKVCY-GLTCLINAFNSYSCSLGSHGVGLCDEEDLFPNQAIKNCK-GK 1421
 Db 789 QTNINE-CASSPLCNQGTCLD-DVAGYKCNKCPLYTGATC---EVVLAFCATSPCSNGSV 843
 Qy 1422 CRLSGLOQPT-CECSSTGTGSDCRELS-----CRGERIDYTGQQQYAAQCTTKVVS 1474
 Db 844 CKSEYDSFSFCVPTGNGQGTCEVDINCKVSPCR-----HG-ASQNTNGSY 891
 Qy 1475 RLECRGGAGGAGC-----CGPLSRKKRYSFECTDG---SSFVDEV-----EKVV 1517
 Db 892 KCLQAGTGTGNCESDIDCRFPNCHN---GGSCDGINAFCDLPGFGQAFCEINE 948
 Qy 1518 C-----GCTCRV 1524
 Db 949 CASNPGQNGACTDCV 964

RESULT 15

A49175
 Notch 9 protein - mouse (fragment)
 N:Alternate names: Notch homolog
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
 C:Accession: A49175; P1570; G32113
 R:Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A:Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety A:Reference number: A49175; MUID:93178563

A:Accession: A49175
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1203 <LAR>
A:Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBI:126158)
C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C:Comment: This protein is one of the neurogenic proteins controlling the decision between
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:143-174/Domain: EGF homology <BG1>
F:482-513/Domain: EGF homology <BG1>
F:560-591/Domain: EGF homology <BGF>
F:674-705/Domain: EGF homology <BG2>
F:712-743/Domain: EGF homology <BG3>
F:836-867/Domain: EGF homology <BG3>

Query Match 8.94; Score 736; DB 2; Length 1203;
Best Local Similarity 22.94; Pred. No. 2.5e-35;
Matches 299; Conservative 111; Mismatches 372; Indels 526; Gaps 61;

434 NPPICDCHLAWLAD-----YLNHPNPIETSGARCT-----SPELRANRIQI 475
Db 65 NP-----CHGALCDZVPLNGQITCTCPQGYKADCTEDVDECAMANSWCEHAGKCVN--- 118
Qy 476 KSKFKRCSTGYD-----RSKL-----SGDCPADLACPEK-----CRCEOTT 512
Db 119 TDGAFHCECLAGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLMPPFGWVCELEV 178
Qy 513 VDC-----SNQKLNKPIEHIPQYTAELRLNNNEFTVLEATGPKPLQRLKRNFSNW- 564
Db 179 NEQCSQPVNNGQCVDKV-----NRQCLCPFGFTGPGVCGIDDCSSTP 223
Qy 565 ----ITDIDEG-AFEGASGVNELLISNRLEN-----VQHMKP-GLSLATYMLRNS 612
Db 224 CLMGAKCIDHFGYECQCATGFTGLICDEN-IDMCDPDCCHGQCQGDIDS----- 273
Qy 613 RITCVDSNGFSLGSLVRLSLDNIITVAPAGFDTLHSLSTLN-----LLANPFMCNCT 667
Db 274 -ITCINQSYNG-----AICSDQI-----DECTSSPLNDAGRIDVNGNQCQ 317
Qy 668 LAWLGEWLKRRITVGNPRQKPYFLKEIPQDVA-----IQDFCTCDGNDNS--CSP--- 719
Db 318 PG-----TSLGNC-----EINFDDCASNPCHBGVCVGINYSVCVSPGFT 358
Qy 720 -----LSRCPIECTCLDVT--VRCSNGLKVLKPGIPRDWTLFDGNQFTLV 765
Db 359 GQRNCNIDECASNPCKRGATCINDVNGFPC-----ICPBG----- 394
Qy 766 PKLSNXYKHLTLDLSNRISTLSNQSFNMTQLLTLISYNLRLCIPPTFGDLSKRL 825
Db 395 PHPFSCTYSQV-----NECLSN-----PCF----- 413
Qy 826 LSLHGNDLSVYPGAPNDLSALSHLAGANPLYCDNNQMLDWSLVSEYKSPGIARCAPG 885
Db 414 -----HGNCCTG-----GLSGYKCLDAGN----- 432
Qy 886 GEMAKDLITLTPSKFTQCGPVDVNLAKCNPLSNPKNDGTCNSDPVTRCTCPGF 945
Db 433 -----VGNCVMDNECLSNPQNGGTGNC-MLVNGYTRCTCKKGF 470
Qy 946 KGQCDVPIHACISNPKKHGGTCH----- 969
Db 471 KGNQCVNIDECASNPCLNQGTCTCFDDVSYTCHCMLPYTGKNCQTVLAPCSNPNCBAV 530
Qy 970 LKGBEG-DGFWICADGFEENCEVNVDDCENNNSTCVGINNTYCLCPPEYTEL 1028
Db 531 CKPAPNFESFSCLCAPGQWQKRCCTVDVDECSKPCMNNGVCHNTQSTVCBCPPGSGMD 590
Qy 1029 CBEKLOFCAQDILNCPQBDKSCILTPKGFCKDCTPGYVGEHCIDIDFQDQCNKNGARCT 1088
Db 591 CREEDINDCL--ANPCQNGSGVDHYNTFSQCHPGFIDGKQCTDMNECLSEPKNGGTCS 648

Qy 1089 DAVNGYTCICPEYSGSLFCE-----FSPFWLPR 1117
Db 649 DYNVSTTCTCPAGFHGVHCENNIDECTESSCFNGGTCVDGINSFCLCPGVGTGPFCLWD 708
Qy 1118 TSPCDFNQCQGAQCIINRINRIPICQCLPGYQGEKCEKLVSV-----NFINKSYLQIPSA 1173
Db 709 INECSNPNCLAGTCTVDGLGTYRCICPLGYTGKNCQTLNWLCSRPFCKNGCTCVQ---EK 765
Qy 1174 VRPO-----TNITLQIAT-----DEDSGILLYKGDKDIHAYELNG 1209
Db 766 ARPHCLPPGMDGATCDVLNWSKAAALQKGVVPEHLCKHSGICINAGNTHHCQCLP--- 822
Qy 1210 RVRASITDTSHPASAIYSVETINDGNFHVVELLALDQSLSLSDVGGNPKIITNLKSQTL 1269
Db 823 ----GT-TGS-----YCEE-----QLDEKAS-----NP-----CQHGTATC 847
Qy 1270 NFDSPLYVGGMPGKSNVASLRQAPGQNGTSFPHGCIRNLYINSELQDPQVPMQTGLPGC 1329
Db 848 ND-----PIGTG-----RCREVPGYQGVN-----CEYEVDEQCNQPCQNG----- 882
Qy 1330 EPCHKKVCARHTCQPSQAGTFCQCGWMPCLDQRTINDCLNCKVCHTCLPIANAFSV 1389
Db 883 -----GTICIDLNV-HFKCSPPFTGRLGLCEBENIDCAGGPHCLNNGGQCVDRIGY 931
Qy 1390 SKCLBGBGVLV--DEEDLEVPQV--AIKCKHKK-----CRLSGLGQP----- 1430
Db 932 TCRCLPGFPGRCBGDINECLSNPSCSBSGLDCVQLKNWYCNICRSAPTRGRCEFTLWC 991
Qy 1431 -----TCESSSYTSDSCDR--EISC-RGERIRDYTQKQGYA 1465
Db 992 PQKPCLMNGSTCAVSNMPPGFCRCPPGFSGARCSQSSQVVKRCRGEQ----- 1039
Qy 1466 ACQTTKYSKSL-----ECRGGCAGGQC-----CGLRSKRKKYKSFEC 1502
Db 1040 -CTHTDSGPRCFCLMPKDCBSCASNPQAGGTCTPQRQPPH-TSRCR 1085

Search completed: January 22, 2001, 12:21:29
Job time: 1766 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:08:19 ; Search time 162.41 Seconds
(without alignments)
303.236 Million cell updates/sec

Title: US-09-540-245A-2
Perfect score: 8316
Sequence: 1 MRGVGMQMLSLGLVLAIL.....SSFYDEKVKVKGCTRCVY 1525

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 88757 seqs, 12294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.39*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3486	41.9	1480	1 SLIT_DROME	P24014 drosophila
2	818	9.8	2703	1 NOTC_DROME	P07207 drosophila
3	759	9.1	2524	1 NOTC_XENLA	P21783 xenopus lae
4	747.5	9.0	2531	1 NTC1_RAT	Q07008 rattus norv
5	745	9.0	2531	1 NTC1_MOUSE	Q01705 mus musculu
6	735.5	8.8	2444	1 NTC1_HUMAN	P46531 homo sapien
7	732.5	8.8	2437	1 NOTC_BRARE	P46530 brachydanio
8	731	8.8	1064	1 FBPI_STRPU	P10079 stronglyloe
9	722	8.7	2318	1 NTC3_MOUSE	Q61982 mus musculu
10	691.5	8.3	1964	1 NTC4_MOUSE	P31695 mus musculu
11	634.5	7.6	2139	1 CRB_DROME	P10040 drosophila
12	625.5	7.5	1408	1 SBRR_DROME	P18168 drosophila
13	606	7.3	1429	1 L112_CAEEL	P14585 caenorhabdi
14	549.5	6.6	570	1 FBPI_STRPU	P49013 stronglyloe
15	543	6.5	1295	1 GLP1_CAEEL	P13508 caenorhabdi
16	535.5	6.4	473	1 FP2_MITGA	Q25464 mytilus gal
17	533.5	6.4	723	1 DLL1_HUMAN	Q00548 homo sapien
18	521.5	6.3	714	1 DLL1_RAT	P97677 rattus norv
19	518.5	6.2	722	1 DLL1_MOUSE	Q61483 mus musculu
20	493.5	5.9	603	1 ALS_MOUSE	P70389 mus musculu
21	491	5.9	833	1 DL_DROME	P10041 drosophila
22	489	5.9	603	1 ALS_RAT	P35859 rattus norv
23	484.5	5.8	605	1 ALS_HUMAN	P35858 homo sapien
24	479.5	5.8	605	1 ALS_PAPPA	O02833 papio papio
25	455.5	5.5	1134	1 CHAO_DROME	P12024 drosophila
26	442.5	5.3	3051	1 YNK3_CAEEL	P34576 caenorhabdi
27	410	4.9	567	1 GPV_MOUSE	O08742 mus musculu
28	403.5	4.9	5147	1 FAT_DROME	P33450 drosophila
29	400	4.8	567	1 GPV_RAT	O08770 rattus norv
30	400	4.8	2871	1 FBNI_BOVIN	P98013 bos taurus
31	397	4.8	2871	1 FBNI_HUMAN	P35555 homo sapien
32	393.5	4.7	2871	1 FBNI_MOUSE	Q61554 mus musculu
33	393	4.7	3707	1 PGM_MOUSE	Q05793 mus musculu

34	390.5	4.7	4393	1 PGM_HUMAN	P98160 homo sapien
35	390	4.7	662	1 GARP_HUMAN	Q14392 homo sapien
36	389.5	4.7	2911	1 FBW2_HUMAN	P35556 homo sapien
37	387.5	4.7	1959	1 AGRI_RAT	P25304 rattus norv
38	386.5	4.6	2907	1 FBW2_MOUSE	Q61555 mus musculu
39	382.5	4.6	560	1 GPV_HUMAN	P40197 homo sapien
40	382	4.6	4590	1 FATH_HUMAN	Q14517 homo sapien
41	376.5	4.5	1097	1 TOLL_DROME	P08953 drosophila
42	375.5	4.5	385	1 DLX_MOUSE	Q09163 mus musculu
43	372	4.5	383	1 DLX_HUMAN	P80370 homo sapien
44	372	4.5	536	1 CRBP_HUMAN	P22792 homo sapien
45	371	4.5	4289	1 TENX_HUMAN	P22105 homo sapien

ALIGNMENTS

RESULT 1
SLIT_DROME
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SLIT PROTEIN PRECURSOR.
GN SLIT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM W.A.
RA MEDLINE-91093665; PubMed-2176636;
RC Rothberg J.M., Jacobs J.R., Goodman C.S., Artavanis-Tsakonas S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL Genes Dev. 4:2169-2187(1990).
CC -1- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -1- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -1- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -1- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -1- SIMILARITY: CONTAINS A C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL: X51959; CAA37910.1; .
DR PIR: A36665; A36665.
DR HSP: P00740; 17XA.
DR FLYBASE: FBgn003425; slit.
DR INTERPRO: IPR000152; .
DR INTERPRO: IPR000359; .
DR INTERPRO: IPR000272; .
DR INTERPRO: IPR000483; .
DR INTERPRO: IPR000561; .
DR INTERPRO: IPR001611; .
DR INTERPRO: IPR001791; .
DR INTERPRO: IPR001801; .
DR PFAM: PF00007; Cys_knot; 1.
DR PFAM: PF00008; EGF; 7.

QY 686 RQKPFYFLKEIPQDAIQDFTCDDGNDNCSPLSCPECTCLODTVRCNSKGLKVL 745
 Db 679 RCGAPSKVRQVQIKDLPHSEFKCSSESS-GLGDGCTPPSCITCTGTVACSRNQKEIP 737
 QY 746 KGIPRDTVELYELDQNOFLVPEK-LSNYKHLTLDLSNRINSTRISNQSNNQTLTLL 804
 Db 738 RGPASTSELYLSNEIQIHYERIRHLSRLTLDLSNQITLISNTYFANLKLSTLL 797
 QY 805 STYRKLIPPTFDGLKSLRLSLHGNDSIVYPGAFNDLSALSHAIKANFLYCDNMQ 864
 Db 798 STYKQLQCHRALSGLNLRVSLHGNRISMLPGSGFEDLSLTHIALGSNFKCDGLK 857
 QY 865 WLSWDVKSSEYKEPGIARCGAPGEMADKLLITPSKTKQCPGVVNLINPCINSPCK 924
 Db 858 WFSWIKLIDYVPGIARCAEPQMDKLLITPSSSVCRGVRVNDILAKCNACFPQPCQ 917
 925 NDGTCNSDPVDFYCTCTPGYKGGQCDVPHACISNPKHGKGTCHLKGESDGFWICAD 984
 918 NQAQVQLPQREYQCLQPGYHSGKCEEMIDACYGPCRNNACTVL-EEGRFSQCAP 975
 QY 985 GFEGECENNVDDC-EDDNCENNSTCVGGNINYLCPPEYTGCEKDLQCAQDLMP 1043
 Db 976 GYTGARCEINIDDCGLKIKQNNATCIDGVESYKCCQPGFSGPCDCTKQCSPEFPC 1035
 QY 1044 QHDSKILPMFGKCDCTPGYVGEHCDIFDDQCKNKGAHCTDAVNGYTCICEPST 1103
 Db 1036 ANGAKCMDFHTYSCDQAGFQGTCTNIDDCNMCGNGCTGVNDQYQCRPDOTT 1095
 QY 1104 GLECFESP-FHWLPPTSCDFDQNGA-QCIVRWBPICQCLPGYQKCEKXILSVN 1159
 Db 1096 GKTCESHMNISMVPTQSPQCNHCKHGVCPQNAQSSDYLCRHPGYTKWCEYLSIS 1155
 QY 1160 PINKSELYIPSAKVRPQNTITLQIATDESSGLLYKDKDHIAYLVGRVSRVSDYGS 1219
 Db 1156 PVHNNSSVLEPLRTPANVTIVSSAQNGLMTDQDAHLAVFLWGRIRVSTDVGN 1215
 QY 1220 HPASALYSVEINQGNPHIVELLADQSLSLSSQGNPKIINLSKSTNLSFDPVGG 1279
 Db 1216 HPVSTISFEMADGKTHAVELLAKANPTLRVDRGLARSINDESSDYLLKILPMFLG 1275
 QY 1280 HPGKSNVASLAPQNGTSFNGCINLINSYELQDQKVPMTGILGPCCECHKVCAH 1339
 Db 1276 LPVDPQAQKANKWILNLSFKGCKHEVINHLKLDVFGNAQKQKITPG-----ALLE 1329
 QY 1340 GTCCPSAGFTCECBGHWG-PLCDQRTNDCPLGNKVHGT-CLP-IVA-FYSCKRL 1394
 1330 GEQEQE-----DDEQDFMDTPIHKEEVPDFLENKRCGRSGVNSNARDGYCKCK 1383
 QY 1395 BGGHGLDCEEDLPNQAICKNGKGLSGGLQPYCECSGTYGDSCEISCRGRI 1454
 Db 1384 HGGGRYTCQGGSTP-----PTVTAI-----TCKRQV 1414
 QY 1455 RDTYQKQGYAAQQTTKVSRLECRGCAGGCGPLSKRRKTSFECTDSSSFVDEYK 1514
 Db 1415 RETTEND-----CBSRQLYKAVGCG-GNQCCKAIVRRKRVNVCNNRKYINLMDI 1469
 QY 1515 VVKGCCTR 1522
 Db 1470 VVKGCCTK 1477

RESULT 2

NOTC_DROME

ID NOTC_DROME STANDARD; PRT: 2703 AA.

AC P07207; P04154;

DT 01-NOV-1986 (Rel. 03, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.

GN N.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86079539; PubMed=3935325;

RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;

RT "Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats.";

RL Cell 43:567-581(1985).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=87064624; PubMed=3097517;

RA Kidd S., Kelley M.R., Young M.W.;

RT "Sequence of the notch locus of Drosophila melanogaster: relationship of the encoded protein to mammalian clotting and growth factors.";

RL Mol. Cell. Biol. 6:3094-3108(1986).

RN [3]

RP SEQUENCE OF 2505-2611 FROM N.A.

RX MEDLINE=85099329; PubMed=2981631;

RA Wharton K.A., Yedvobnick B., Pinerly V.G., Artavanis-Tsakonas S.;

RT "opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster.";

RL Cell 40:55-62(1985).

RN [4]

RP SEQUENCE OF 1-8 FROM N.A.

RX MEDLINE=87257846; PubMed=3037327;

RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;

RT "Restriction of P-element insertions at the Notch locus of Drosophila melanogaster.";

RL Mol. Cell. Biol. 7:1545-1548(1987).

RN [5]

RP REVIEW.

RA Harris W.A.;

RT "Many cell types specified by Notch function.";

RL Curr. Biol. 1:120-122(1991).

CC [1]

CC FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM.

CC [1]- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC [1]- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO

CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS

CC DEVELOPMENT IN INSECTS. THIS PROCESS IS UNDER CONTROL OF THE

CC NEUROGENIC GENES.

CC [1]- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

CC [1]- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.

CC [1]- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.

CC [1]- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC [3]

DR EMBL; M16152; AAB59220.1; .

DR EMBL; M16153; AAB59220.1; JOINED.

DR EMBL; M16149; AAB59220.1; JOINED.

DR EMBL; M16150; AAB59220.1; JOINED.

DR EMBL; M16151; AAB59220.1; JOINED.

DR EMBL; K03508; AAA28725.1; .

DR EMBL; M13689; AAA28725.1; JOINED.

DR EMBL; K03507; AAA28725.1; JOINED.

DR EMBL; M12175; AAA74496.1; .

DR EMBL; M16025; AAA28726.1; .

DR PIR; A24420; A24420.

DR PIR; A24768; A24768.

DR PIR; A05267; A05267.

DR HSSP; P00740; IIXA.

DR FLXBASE; FB900004647; N.

DR INTERPRO; IP0000152; .

FT	DISULFID	166	175	BY	SIMILARITY
FT	DISULFID	181	192	BY	SIMILARITY
FT	DISULFID	186	203	BY	SIMILARITY
FT	DISULFID	205	214	BY	SIMILARITY
FT	DISULFID	221	232	BY	SIMILARITY
FT	DISULFID	226	241	BY	SIMILARITY
FT	DISULFID	243	252	BY	SIMILARITY
FT	DISULFID	259	270	BY	SIMILARITY
FT	DISULFID	264	279	BY	SIMILARITY
FT	DISULFID	281	290	BY	SIMILARITY
FT	DISULFID	297	308	BY	SIMILARITY
FT	DISULFID	302	317	BY	SIMILARITY
FT	DISULFID	319	328	BY	SIMILARITY
FT	DISULFID	335	349	BY	SIMILARITY
FT	DISULFID	343	358	BY	SIMILARITY
FT	DISULFID	360	369	BY	SIMILARITY
FT	DISULFID	376	387	BY	SIMILARITY
FT	DISULFID	381	396	BY	SIMILARITY
FT	DISULFID	398	407	BY	SIMILARITY
FT	DISULFID	413	424	BY	SIMILARITY
FT	DISULFID	418	435	BY	SIMILARITY
FT	DISULFID	437	446	BY	SIMILARITY
FT	DISULFID	453	465	BY	SIMILARITY
FT	DISULFID	459	474	BY	SIMILARITY
FT	DISULFID	476	485	BY	SIMILARITY
FT	DISULFID	492	503	BY	SIMILARITY
FT	DISULFID	497	512	BY	SIMILARITY
FT	DISULFID	514	523	BY	SIMILARITY
FT	DISULFID	530	541	BY	SIMILARITY
FT	DISULFID	535	550	BY	SIMILARITY
FT	DISULFID	552	561	BY	SIMILARITY
FT	DISULFID	568	579	BY	SIMILARITY
FT	DISULFID	573	588	BY	SIMILARITY
FT	DISULFID	590	599	BY	SIMILARITY
FT	DISULFID	606	616	BY	SIMILARITY
FT	DISULFID	611	625	BY	SIMILARITY
FT	DISULFID	627	636	BY	SIMILARITY
FT	DISULFID	643	654	BY	SIMILARITY
FT	DISULFID	648	663	BY	SIMILARITY
FT	DISULFID	665	674	BY	SIMILARITY
FT	DISULFID	681	692	BY	SIMILARITY
FT	DISULFID	686	701	BY	SIMILARITY
FT	DISULFID	703	712	BY	SIMILARITY
FT	DISULFID	719	730	BY	SIMILARITY
FT	DISULFID	724	739	BY	SIMILARITY

Query Match 9.8%; Score 818; DB 1; Length 2703;
Best Local Similarity 29.3%; Pred. No. 3.7e+43;
Matches 262; Conservative 93; Mismatches 336; Indels 202; Gaps

Qy	707	TCDDGNDNDSQSLPSRCTPCTCT---CLDTVVRKSNKGLKPLKPGIDPVTELYLDGNQFT	763
Db	307	TCIGDISDYTC-----RCPNFTGRCFGQDDVDEACGRDHPVQNGATCTNH-----	353
Qy	764	LVPKELSNKMKLTL-----IDLNSNRISTNSQFSNQTQLLTLLSYNLRCLPPTPTFD	818
Db	354	-----GYSTICIGNWAGLDLNNWIDCKQACQFYGAT-----CT-----	389
Qy	819	GLKSLRLLSLNGKNSIVSPPEAFNDSLALSHL-AIGANPLYCD-CNMQLSD---WW	870
Db	390	GVGSFYCQCTGK-----TGLLCHLDADCTSNCHADAIDCTSPINGYACSC	437
Qy	871	KSEYK-----EPGIARC-AGPGEMADKLLTTPSKAKFTCQ---GP-VDVNMLKCNPL	919
Db	438	ATIIYKGVDCSDIDEDCQSGPCHENGICVYTPGYSRCNCQQGTGRCPTNI---WECE	493
Qy	920	SNPCENDGCTGSDPVDVYRCTCPYFGKQDCDVPFHACINSNPKHGGTCHLKEGDEGFW	979
Db	494	SHPCQNBGSGLDLPGTF-RCVMPFGFTQCEIDIDEDCQSNPCINDGTGCHCK--INGFR	549
Qy	980	CICADGFBEGECENVDWDCEDNCCNSENVDGINNYTCLPPTPTGELCEEKLDQCAQD	1033
Db	550	CSGALCGFTGARCOINIDDCQSPCPNRNGICHSIAGYSCEBCPGYGTGSCSEINDC--D	607

Query Match 9.88; Score 818; DB 1; Length 2703;
Best Local Similarity 29.38; Pred. No. 3.7e-43;
Matches 262; Conservative 93; Mismatches 336; Indels 202; Gaps 50;

```

QY 707 TCGDNDGDSNFSPLSRPCTCT---CLDTVRSKMGKLVKPGIDPDTVELYLDGQPT 763
Db 307 TCGIDISDITCT---RCPPNFTRGFCQDVEDAGQORDHVPQMGAPCTNTH----- 353

QY 764 LVPKLSNTKLKLTL---IDLNSNRITSLNGSFSNMTQLLTILSYNLRCLPPTPTFD 818
Db 354 ---GTSYCIQYWGALDLSNNTDCKQAQCFYAT-----CT---D 389

QY 819 GLKSLRLSLGKGNDSIVPPEAFNDLSLSHL- AIGANPLYCD- CNMQWLS-----WY 870
Db 390 GVGSTFYCQCTGK-----TGLLCHLDATCSNPCHDAICDTPSINGSYACSC 437

QY 871 KSEYK---EPGIARC-AGPGEMADKLITLTPSKPTCTQ---GP-VQVNLKACNPL 919
Db 438 APTAGVDCSDIEDDQSGSPCHNGICVWTPGYSKNCSGQGFGRCTNT---WEE 493

QY 920 SNPGKAGDSCSDPWFYKCTPGYKGCQDCVPHACISNPAGKGTGCHTAAEGEED 979
Db 494 SHPCWNESSDLDGPTF-KCVCMFPGTGQCIIECDQSCNPLNDGCTN---INGFP 549

QY 1033 CACDAGPGEENCVEVNDGNDENNSTNCSTGGINNYCLCPPEYTGECHEKRLDCAQD 1033
Db 550 CSCAAGFYGARCOINIDDCOSCPNRNGCHDSIAGYSCBPPIYGTGSCENINDC-D 607

```


Qy 1040 LNPQHDHSCILTPGKFCDCPTGVYGHCDIDFDQDQNKCKNGARCTDAVNGTCTICP 1099
 Db 608 SNPC-HRGKICIDVWFSKCLDQPGYTGTCQKQINCEBSPNCPQDGRGQDQVSGTTCQ 666
 Qy 1100 BGYSGLPCEFPFVLPRTSPDNFDCQNGAQICVINEPIQCLPQYQGEKXLSVSN 1159
 Db 667 AGTSKCNFVN-----VNECHSNPCNGATCDGINSYKQCVPGFTGQCEX----- 714
 Qy 1160 FINKESYLQIPSAKVRPQTNIQLIATDEDSGLLYKGDKHIAVELTRGVRASTDTGS 1219
 Db 715 -NVDCISSEPCA-----NMGVCIQVNG--YK-----CECPG---PYD--A 748
 Qy 1220 HPASAI-----YSVETINDGNFHVELLADQSLSDVSGNPKII-----THLSQSTLNF 1272
 Db 749 HCLSDVDECAENPCNDEGRCE-----DGINEFICHPPGYTGKRCLEID 793
 1273 -----SPLVYG-MPGKSNVSLRQAPQNG-----TSFHCGRINLIY-----SELQD- 1317
 794 ECSSNPQGGGTCITKILNAPSCQCMFGYTGKCTCINDVCPNPGNGGTCIKRWYMKIC 853
 Qy 1318 -----VYPMGTGLPGCE-----PCHKKVCAN-GTCQPSQ-AGTCEBQGMPLCDQRTND 1369
 Db 854 WKVPY-TG--RDCESKMDPCANBCKNEAKCTPSSNLDPSCTCKLGTGRCDDEIDE 910
 Qy 1370 PCLNGCNCVWG-TCLPINAFTSYCKCLRGHGVGLDDEEDLNPQCAIKCKHCKRLSLGL 1428
 Db 911 CSLSSPCNGASCLANVPG-SYRCLCTKYEGRDAINTD-----DCASPCQNGGTCLDGL 966
 Qy 1429 QPVCBSSGTYGDSQDREIS-----CR-GERIDRYQKQ-----GYAA--CQTKKVS 1474
 Db 967 DYSLCLVDGFGDKRCHETDINELCSQPCQNGATCSQVNSYTCPLGFSINQVND-- 1024
 Qy 1475 RLBC-ROGCA-QGQCCPLASKRRKYSPECTGSSVDFEVKXVGCQCTKVS 1535
 Db 1025 -DCTBSSCLAGGSGDING-----YMCSLAGYSANQYKLNKRCDSNPLN 1071

RESULT 3
 NOTCH_XENLA
 ID NOTCH_XENLA STANDARD: PRT: 2524 AA.
 C P21783;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (NOTCH PROTEIN).
 KOTCH.
 Xenopus laevis (African clawed frog).
 CC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Duteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipeloidae; Pipidae;
 CC Xenopodinae; Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90385285; PubMed=2402639;
 RA Coffman C., Harris W., Kintner C.;
 RT "Xotch, the Xenopus homolog of Drosophila notch";
 RL Science 249:1438-1441(1990).
 RN [2]
 RP REVISIONS TO 1759-1782.
 RA Kintner C.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M33874; AAB02039.1; .
 DR PIR: A35844; A35844.
 DR HSSP: P00740; 1IXA.
 DR INTERPRO: IP8000152; .
 DR INTERPRO: IP8000561; .
 DR INTERPRO: IP8000800; .
 DR INTERPRO: IP8001438; .
 DR INTERPRO: IP8001881; .
 DR INTERPRO: IP8002110; .
 DR PFAM: PF00008; EGF; 36.
 DR PFAM: PF00023; ank; 6.
 DR PFAM: PF00066; notch; 3.
 DR PRINTS: PR00010; EGFBL00D.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 23.
 DR PROSITE: PS00022; EGF_1; 34.
 DR PROSITE: PS01186; EGF_2; 29.
 DR PROSITE: PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 19
 FT CHAIN 20 2524
 FT DOMAIN 20 1728
 FT TRANSMEM 1729 1750
 FT DOMAIN 1751 2524
 FT DOMAIN 20 57
 FT DOMAIN 58 99
 FT DOMAIN 102 140
 FT DOMAIN 141 177
 FT DOMAIN 179 215
 FT DOMAIN 217 254
 FT DOMAIN 256 292
 FT DOMAIN 294 332
 FT DOMAIN 334 370
 FT DOMAIN 371 409
 FT DOMAIN 411 449
 FT DOMAIN 451 487
 FT DOMAIN 489 525
 FT DOMAIN 527 563
 FT DOMAIN 565 600
 FT DOMAIN 602 638
 FT DOMAIN 640 675
 FT DOMAIN 677 713
 FT DOMAIN 715 750
 FT DOMAIN 752 788
 FT DOMAIN 790 826
 FT DOMAIN 828 866
 FT DOMAIN 868 904
 FT DOMAIN 906 942
 FT DOMAIN 944 980
 FT DOMAIN 982 1018
 FT DOMAIN 1020 1056
 FT DOMAIN 1058 1094
 FT DOMAIN 1096 1142
 FT DOMAIN 1144 1180
 FT DOMAIN 1182 1218
 FT DOMAIN 1220 1264
 FT DOMAIN 1266 1304
 FT DOMAIN 1306 1346
 FT DOMAIN 1347 1383
 FT DOMAIN 1385 1424
 FT DOMAIN 1441 1560
 FT REPEAT 1441 1478
 FT REPEAT 1479 1520
 FT REPEAT 1521 1560
 FT DOMAIN 1871 2083
 FT DISULFID 22 35
 FT DISULFID 29 45
 FT DISULFID 47 56
 POTENTIAL.
 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 6.
 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 10.
 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 17.
 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 22.
 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 26.
 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 28.
 EGF-LIKE 29.
 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 33.
 EGF-LIKE 34.
 EGF-LIKE 35.
 EGF-LIKE 36.
 3 X LIN/NOTCH REPEATS.
 LIN/NOTCH 1.
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 6 X ANK MOTIF REPEATS.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT DISULFID 62 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 128 BY SIMILARITY.
 FT DISULFID 130 139 BY SIMILARITY.
 FT DISULFID 145 156 BY SIMILARITY.
 FT DISULFID 150 165 BY SIMILARITY.
 FT DISULFID 167 176 BY SIMILARITY.
 FT DISULFID 183 194 BY SIMILARITY.
 FT DISULFID 188 203 BY SIMILARITY.
 FT DISULFID 205 214 BY SIMILARITY.
 FT DISULFID 221 232 BY SIMILARITY.
 FT DISULFID 226 242 BY SIMILARITY.
 FT DISULFID 244 253 BY SIMILARITY.
 FT DISULFID 260 271 BY SIMILARITY.
 FT DISULFID 265 280 BY SIMILARITY.
 FT DISULFID 282 291 BY SIMILARITY.
 FT DISULFID 298 311 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 338 349 BY SIMILARITY.
 FT DISULFID 343 358 BY SIMILARITY.
 FT DISULFID 360 369 BY SIMILARITY.
 FT DISULFID 375 386 BY SIMILARITY.
 FT DISULFID 380 397 BY SIMILARITY.
 FT DISULFID 399 408 BY SIMILARITY.
 FT DISULFID 415 428 BY SIMILARITY.
 FT DISULFID 422 437 BY SIMILARITY.
 FT DISULFID 439 448 BY SIMILARITY.
 FT DISULFID 455 466 BY SIMILARITY.
 FT DISULFID 460 475 BY SIMILARITY.
 FT DISULFID 477 486 BY SIMILARITY.
 FT DISULFID 493 504 BY SIMILARITY.
 FT DISULFID 498 513 BY SIMILARITY.
 FT DISULFID 515 524 BY SIMILARITY.
 FT DISULFID 531 542 BY SIMILARITY.
 FT DISULFID 536 551 BY SIMILARITY.
 FT DISULFID 553 562 BY SIMILARITY.
 FT DISULFID 569 579 BY SIMILARITY.
 FT DISULFID 574 588 BY SIMILARITY.
 FT DISULFID 590 599 BY SIMILARITY.
 FT DISULFID 606 617 BY SIMILARITY.
 FT DISULFID 611 626 BY SIMILARITY.
 FT DISULFID 628 637 BY SIMILARITY.
 FT DISULFID 644 654 BY SIMILARITY.
 FT DISULFID 649 663 BY SIMILARITY.
 FT DISULFID 665 674 BY SIMILARITY.
 FT DISULFID 681 692 BY SIMILARITY.
 FT DISULFID 686 701 BY SIMILARITY.
 FT DISULFID 703 712 BY SIMILARITY.
 FT DISULFID 719 729 BY SIMILARITY.
 FT DISULFID 724 738 BY SIMILARITY.
 FT DISULFID 740 749 BY SIMILARITY.
 FT DISULFID 756 767 BY SIMILARITY.
 FT DISULFID 761 776 BY SIMILARITY.
 FT DISULFID 778 787 BY SIMILARITY.
 FT DISULFID 794 805 BY SIMILARITY.
 FT DISULFID 799 814 BY SIMILARITY.
 FT DISULFID 816 825 BY SIMILARITY.
 FT DISULFID 832 843 BY SIMILARITY.
 FT DISULFID 837 854 BY SIMILARITY.
 FT DISULFID 856 865 BY SIMILARITY.
 FT DISULFID 872 883 BY SIMILARITY.
 FT DISULFID 877 892 BY SIMILARITY.
 FT DISULFID 894 903 BY SIMILARITY.
 FT DISULFID 910 921 BY SIMILARITY.
 FT DISULFID 915 930 BY SIMILARITY.
 FT DISULFID 932 941 BY SIMILARITY.
 FT DISULFID 966 997 BY SIMILARITY.
 FT DISULFID 991 1006 BY SIMILARITY.
 FT DISULFID 1008 1017 BY SIMILARITY.
 FT DISULFID 1024 1035 BY SIMILARITY.

FT DISULFID 1029 1044 BY SIMILARITY.
 FT DISULFID 1046 1055 BY SIMILARITY.
 FT DISULFID 1062 1073 BY SIMILARITY.
 FT DISULFID 1067 1082 BY SIMILARITY.
 FT DISULFID 1084 1093 BY SIMILARITY.
 FT DISULFID 1100 1121 BY SIMILARITY.
 FT DISULFID 1115 1130 BY SIMILARITY.
 FT DISULFID 1132 1141 BY SIMILARITY.
 FT DISULFID 1148 1159 BY SIMILARITY.
 FT DISULFID 1153 1168 BY SIMILARITY.
 FT DISULFID 1170 1179 BY SIMILARITY.
 FT DISULFID 1186 1197 BY SIMILARITY.
 FT DISULFID 1191 1206 BY SIMILARITY.
 FT DISULFID 1208 1217 BY SIMILARITY.
 FT DISULFID 1224 1243 BY SIMILARITY.
 FT DISULFID 1237 1252 BY SIMILARITY.
 FT DISULFID 1254 1263 BY SIMILARITY.
 FT DISULFID 1270 1283 BY SIMILARITY.
 FT DISULFID 1275 1292 BY SIMILARITY.
 FT DISULFID 1294 1303 BY SIMILARITY.
 FT DISULFID 1310 1321 BY SIMILARITY.

Query Match 9.14; Score 759; DB 1; Length 3524;
 Best Local Similarity 24.34; Pred. No. 1.7e-39;
 Matches 245; Conservative 114; Mismatches 564; Indels 286; Gaps 48;

Qy 644 GADTILBSLSTLNLANPFWNCYLAHLGEMLAKKRVITGMP-----RCQNPFFLKEIPI 698
 Db 115 GTCELANSTV-----ETKCRCPGWGDSQOQAPCASNPANGKCLPFETQICK 166
 Qy 699 QDAVITQTC-DDGND--DNSCSPLSRCPTE-----CYCLDTVVRCSNKLGLPKGIP 750
 Db 167 CPPGFGATCAQDINCSQNPCKNGGQINFEFSYCTCQN-----RFTGR 212
 Qy 751 DVEYLIDGNGFTLVKELSNYKHLTLDLSNWRISTLSNQFSNMTQL----- 800
 Db 213 NCEPTEVPCN-----PSPCLNGTCTQDSDTSIDTCLGFGSGNCEENIDDCPSNCRN 267
 Qy 801 ---TLISYNYLRKIPPTFDGLSKLLSLNGNDISVVPFGAFNDSLSHAIGANPLY 858
 Db 268 GGTGVGVYINQCPFWGTG-----QTCEBDVDCQLPMACQN--GTCNHTGGS--YN 319
 Qy 859 CDDNMQLMSWVRSYKEPIKARCAGGEMADKLLTTPSKTKQCPQWD-VNIAKCN 916
 Db 320 CVCVNGWGTGDCSENIDCANACHSGATCHDRV-----ASFTCECPHRTGLLCHDN 373
 Qy 917 PCLSNPKNDGTCNSDPVDFTF-CITCPGFGKGQDVPFHACI--SNPKHGSGTCHLHG 973
 Db 374 ACISNPKNGSCNDTPVNGKATCTCPGPTGACNNDVDELSGANPCGSGRGTNLG 433
 Qy 974 HEDGPGKICADGFEGBECNVDVDCDCHDNNSTCVGGINNYTCLCPPTGELCEKL 1033
 Db 434 ---SPFCNPGGYAGRCEDINVECLSNPCNDSTCLQIGEPFCICMHPGELVYCETNI 490
 Qy 1034 DFCQADLPNQHSDKSLITPKFGKCDCTPGYVGHCDIDFDCDCKNCKGAMCTDAWG 1093
 Db 491 DEAS--NPLCENKGLDKNEPRDCDPTGFGNLCOHQDFECTSTCKNGAKCLDGNPS 548
 Qy 1094 YTCICPEGTSGLFCEFSPPWLVPTSPCNDFNQGAQCIVRIPEIQCLPGTQGEKCE 1153
 Db 549 YTCQCTBGTGHCQDINECIP--DPCHYCTCKDGIATFT-----CLCRPGTGLGLCD 600
 Qy 1154 KLVSYFNKESILQIPSAKVRPQTNIQTATDEDSGILLY--KGDKHIAVELYGRV 1211
 Db 601 -----NDINE-----CLSKPLNGQG--CTDRENGYCTCPKG----- 631
 Qy 1212 RASDTGSHSPASITSVETINDGNEHIVELLADQSLSLSVGGNPKIITNLSKSTLNF 1271
 Db 632 ---TTG-----VNCET-----KIDDCASNLCDNG--KCIDKIDGYECT-- 664
 Qy 1272 DPLSYVGMPKSNWASLQAQPGQNTSF-----HGCINRLYINSELQ 1314
 Db 665 CBPGTIGKL-CNININECDNSCPNCRGGTCKDQINGPTCVCPDGHDMNCL-----SEVN 717

AC NCBI.RAT STANDARD; PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 12-JUL-1997 (Rel. 38, Last sequence update)
DT 12-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridea; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SCHWANN CELL;
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of *Drosophila* Notch expressed during mammalian
development.";
RL Development 113:199-205 (1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
SU Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC OF TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC ADULT.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC ~~~~~
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annotation>
CC or send an email to license@isb-sib.ch).
CC ~~~~~
DR EMBL: X57405; CAA00667.1; -.
DR HSP: P00740; 11NM.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000800; -.
DR INTERPRO: IPR001438; -.
DR INTERPRO: IPR001881; -.
DR INTERPRO: IPR002049; -.

INTERPRO: IPR002110; -
DR PFAM: PF00008; BGF_36.
DR PFAM: PF00203; ank; 6.
DR PFAM: PF00066; notch; 3.
DR PRINTS: PRO0010; BGFLOOD.
DR PRINTS: PRO0011; BGFAMININ.
DR PROSITE: PS00088; ANK_REPEAT; 4.
DR PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASK_HYDROXYL; 22.
DR PROSITE: PS00022; BGF_L1; 35.
DR PROSITE: PS01186; BGF_2; 2.
DR PROSITE: PS01187; BGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; BGF-like domain;
Transmembrane; Signal; glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).
FT TRANSDOM 1724 1746 POTENTIAL.
FT DOMAIN 1747: 2531 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 BGF-LIKE 1.
FT DOMAIN 59 99 BGF-LIKE 2.
FT DOMAIN 102 139 BGF-LIKE 3.
FT DOMAIN 140 176 BGF-LIKE 4.
FT DOMAIN 178 216 BGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 218 255 BGF-LIKE 6.
FT DOMAIN 257 293 BGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 295 333 BGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 335 371 BGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 410 BGF-LIKE 10.
FT DOMAIN 412 450 BGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 452 488 BGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 BGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 BGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 BGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 BGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 BGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 BGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 716 751 BGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 BGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 BGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 BGF-LIKE 22.
FT DOMAIN 869 905 BGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 BGF-LIKE 24.
FT DOMAIN 945 981 BGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 BGF-LIKE 26.
FT DOMAIN 1021 1057 BGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 BGF-LIKE 28.
FT DOMAIN 1097 1143 BGF-LIKE 29.
FT DOMAIN 1145 1181 BGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 BGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 BGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 BGF-LIKE 33.
FT DOMAIN 1307 1346 BGF-LIKE 34.
FT DOMAIN 1348 1384 BGF-LIKE 35.
FT DOMAIN 1387 1426 BGF-LIKE 36.
FT DOMAIN 1449 1462 CYS-RICH.
FT DOMAIN 1865 2076 6 X ANK MOTIF REPEATS.
FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2076 ANK MOTIF 6.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.

Oy 1516 VKC-----GTCRCV 1524
 :| | | |
 Db 947 NCAATNPQNGANTCDV 964

RESULT 5
 NTC1_MOUSE
 ID NTC1_MOUSE STANDARD: PRT: 2531 AA.
 AC Q01705:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo P., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RW [2]
 RP SEQUENCE OF 1551-2170 FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo P., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan N.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: L11886; CAA77941.1; -.
 DR HSSP: P00740; 11XA.
 DR MGO: MGI:97363; NOTCH1.
 DR INTERPRO: IP8000152; -.
 DR INTERPRO: IP8000561; -.
 DR INTERPRO: IP8000800; -.
 DR INTERPRO: IP8001438; -.
 DR INTERPRO: IP8001881; -.
 DR INTERPRO: IP8002110; -.
 DR PFAM: PF00008; EGF; 35.
 DR PFAM: PF00023; ank; 6.
 DR PFAM: PF00066; notch; 3.
 DR PRINTS: PR0010; EGFLOOD.
 DR PROSITE: PS50086; ANK_REPEAT; 2.
 DR PROSITE: PS50237; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50010; ASX_HYDROLYS; 22.
 DR PROSITE: PS50022; EGF_1; 34.
 DR PROSITE: PS01186; EGF_2; 27.
 DR PROSITE: PS01187; EGF_C; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.

FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1726 1746 POTENTIAL.
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 1426 36 X EGF-TYPE REPEATS.
 FT DOMAIN 20 58 EGF-LIKE 1.
 FT DOMAIN 59 99 EGF-LIKE 2.
 FT DOMAIN 102 139 EGF-LIKE 3.
 FT DOMAIN 140 176 EGF-LIKE 4.
 FT DOMAIN 178 216 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 218 255 EGF-LIKE 6.
 FT DOMAIN 257 293 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 295 333 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 335 371 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 372 410 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 412 450 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 907 943 EGF-LIKE 24.
 FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 983 1019 EGF-LIKE 26.
 FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1059 1095 EGF-LIKE 28.
 FT DOMAIN 1097 1143 EGF-LIKE 29.
 FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1267 1305 EGF-LIKE 33.
 FT DOMAIN 1307 1346 EGF-LIKE 34.
 FT DOMAIN 1348 1384 EGF-LIKE 35.
 FT DOMAIN 1387 1426 EGF-LIKE 36.
 FT DOMAIN 1449 1462 CYS-RICH.
 FT DOMAIN 1445 1562 3 X LIN/NOTCH REPEATS.
 FT REPEAT 1445 1480 LIN/NOTCH 1.
 FT REPEAT 1481 1522 LIN/NOTCH 2.
 FT REPEAT 1523 1562 LIN/NOTCH 3.
 FT DOMAIN 1865 2075 6 X ANK MOTIF REPEATS.
 FT REPEAT 1865 1910 ANK MOTIF 1.
 FT REPEAT 1912 1942 ANK MOTIF 2.
 FT REPEAT 1944 1975 ANK MOTIF 3.
 FT REPEAT 1978 2009 ANK MOTIF 4.
 FT REPEAT 2011 2042 ANK MOTIF 5.
 FT REPEAT 2044 2075 ANK MOTIF 6.
 FT DISULFID 24 37 BY SIMILARITY.
 FT DISULFID 31 46 BY SIMILARITY.
 FT DISULFID 63 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 127 BY SIMILARITY.
 FT DISULFID 129 138 BY SIMILARITY.
 FT DISULFID 144 155 BY SIMILARITY.
 FT DISULFID 149 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 182 195 BY SIMILARITY.
 FT DISULFID 189 204 BY SIMILARITY.
 FT DISULFID 206 215 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.
 FT DISULFID 266 281 BY SIMILARITY.

Qy 662 FNCNCLAWLGEWLKKRIVTGNPRQCYPFLKSIPIQDAVDFTCDGMDNSCSPLS 721
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 202 YRCACCATHTG-----PHCELTPY-----VPCSPSPQMG-----ATCRPTG 237

Qy 722 RCPTECTCLDTVV-----RCSNGLKVLPGKI-----PRDVTLEYL--D 758

Db 238 DTHTEACILPFAQNCSENVDCPGNKNKGACV--DGWYTNCRCPPEVYGCTCED 295

Qy 759 GHOFTVFLKELSNKHLTLIDLNSNRISTGNSFSNQLTLILSYNRLACIPPTFD 818

Db 296 VDSQCMLPNAQCN-----AGTCHN----- 314

Qy 819 GLSLRLSLSLGNSIDVYPGFAFNOLSAHSLAIGANPLYCDDCNQWLSDWKSYSEKPE 878

Db 315 -----THGYN-----CVCVNGWTGEDCSINDDCA 340

Qy 879 IARCAGFGEMADKLLITPYSKFTTCQGPVD-VHILAK-NFLCNSPNCKNDGNSDPDF 93

Db 341 SAAFCPAGTCHDRV-----ASTPCECPHGTGLVLRHACISPCNCGSSGKPNWPG 394

Qy 937 YR-CICPGFPGK-----QD-----COWPH 955

Db 395 KRITCPSGTSPACSDQVDECLGAKNCHAGKCLMTLGSFEGCQQLQGTGCGCIDW 454

Qy 956 ACISNPCKNHGCTCHLAGEEDGFWCIDGFBEGECENVWDCEDCENNSCTVDGNIW 1015

Db 455 ECISNPQDANCLDQIGE-----PQCPCPBGVGEIWDGACSPCLNAGWCKHDE 511

Qy 1016 YTLCPCPPTTGELCELDQFCAQDLNPQKDSKILPFGKPCGCTPGTVGEHCIDFDD 1075

Db 512 PQCCPCPFGNHLGQLDWDGAS--TPCKNAKCLAGSTVTCVTBGTGTCCCEVDE 569

Qy 1076 QDQNKNCNGHACTDAVNGYTICPEGYSGLCFSPFVMLPSTPCDNPDCQAGCIVR 1135

Db 570 CDQDPCHYGS--CKDGYATFLCQLQGTGHC-----TINIECHSGPCRHGTCQDR 621

Qy 1136 INEPIQLCPQTGKECKEKLVSYNFINESKYLISAKVRYQTNLTQIATDEDSGLLY 1195

Db 622 DNSYLCLLGTITGPNCE-----INDDCASNPC-----DSGTCLD 657

Qy 1196 RKGKDHIAVEL-YRGVRYASTDQGHSPASITSYVETINDG-----NFHIVELL 1244

Db 658 KIDGFCACEPGETGSM-CVNWIDECAGSPCHMGQCEDLGFTCRCPGTH----- 709

Qy 1245 DQSLSLSYVGNGPKIITNLKSQSTLNDSPSLYVGMGPKNVASLRQAQMGWTSFBCI 1304

Db 710 -----DP-----TCLSEYVENCN-SNFCIHGACRDLGALGYKCCAPGSGT----- 748

Qy 1305 RNLYINSELQDPQRYPMQGLGPCRCCHKVKAH-GTQCPSQAGFTCECQBGNGMPLC 1363

Db 749 -NCDLIN-----NKESENPCVNGTCKDMT-SGVTCYCRBGSGPNC 788

Qy 1364 DQRTNDPLCKMKYV-GTCLPINAFYSKRCLEHGSGVLCDEEDLFPNQAIKNC-GR 1421

Db 789 QTINIE-CASHPICLNGTICD-DVAGKYCNCPPLYTGATC-----EVLAPCATSPKNSGV 843

Qy 1422 CRLSGLOQPY-CEBSSGYTGDSCDREIS-----CRGERIRDYTQQQGTAACTTKRVS 1474

Db 844 CKSEDEYESFSQVCTPGMGQGTCEVDINECKSPCR-----HG-ASQONTNGSY 891

Qy 1475 RLBCRGAGGACG-----CGPLSRKARKYSFECTDG--SSFVDEY-----EKXVV 1517

Db 892 RLCCQAGYTGNCESDIDDCRPNCHN-----GSGCTDGINATPCDLPGFGAPCEDINE 948

Qy 1518 C-----GCTRCV 1524

Db 949 CASNPQNGANCTDVCY 964

```

RESULT      6
NTCL_HUMAN
ID   NTCL_HUMAN      STANDARD;          PRT;   2444 AA.
AC   P46531;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE   ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN   NOTCEL OR TAN1.

```

OS Homo sapiens (Human).
OC Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91347357; PubMed-1831692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms";
RL Cell 66:649-661(1991).
CC -1- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M73990; AAA60614.1; ..
DR HSSP: P00740; LIXA.
DR MIM: 190198; ..
DR INTERPRO: IPR000152; ..
DR INTERPRO: IPR000561; ..
DR INTERPRO: IPR000800; ..
DR INTERPRO: IPR001881; ..
DR INTERPRO: IPR002110; ..
DR PFM: PFM00008; EGF; 36.
DR PFM: PFM00023; ank; 6.
DR PFM: PFM00066; notch; 3.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50010; ASX_HYDROXYL; 20.
DR PROSITE: PS50022; EGF_1; 34.
DR PROSITE: PS01185; EGF_2; 26.
DR PROSITE: PS01187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 >2444 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT DOMAIN 19 1736 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1737 1757 POTENTIAL.
FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 EGF-LIKE 1.
FT DOMAIN 59 99 EGF-LIKE 2.
FT DOMAIN 102 139 EGF-LIKE 3.
FT DOMAIN 140 176 EGF-LIKE 4.
FT DOMAIN 178 216 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 218 255 EGF-LIKE 6.
FT DOMAIN 257 293 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 295 333 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 335 371 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 410 EGF-LIKE 10.
FT DOMAIN 412 450 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 20.
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 868 EGF-LIKE 22.
FT DOMAIN 870 906 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 908 944 EGF-LIKE 24.
FT DOMAIN 946 982 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 984 1020 EGF-LIKE 26.
FT DOMAIN 1022 1058 EGF-LIKE 27.
FT DOMAIN 1060 1096 EGF-LIKE 28.
FT DOMAIN 1098 1144 EGF-LIKE 29.
FT DOMAIN 1146 1182 EGF-LIKE 30.
FT DOMAIN 1184 1220 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1222 1266 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1268 1306 EGF-LIKE 33.
FT DOMAIN 1308 1347 EGF-LIKE 34.
FT DOMAIN 1349 1385 EGF-LIKE 35.
FT DOMAIN 1388 1427 EGF-LIKE 36.
FT DOMAIN 1446 1563 3 X LIN/NOTCH REPEATS.
FT REPEAT 1446 1481 LIN/NOTCH 1.
FT REPEAT 1482 1523 LIN/NOTCH 2.
FT REPEAT 1524 1563 LIN/NOTCH 3.
FT DOMAIN 1876 2087 6 X ANK MOTIF REPEATS.
FT REPEAT 1876 1921 ANK MOTIF 1.
FT REPEAT 1923 1954 ANK MOTIF 2.
FT REPEAT 1956 1987 ANK MOTIF 3.
FT REPEAT 1990 2021 ANK MOTIF 4.
FT REPEAT 2023 2054 ANK MOTIF 5.
FT REPEAT 2056 2087 ANK MOTIF 6.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1562 1565 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.

CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
 CC BRAIN AND HEAD REGIONS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
 CC OR SEND AN EMAIL TO license@isb-sib.ch).
 CC -----
 DR EMBL: X59088; CAA48831.1; .
 DR HSSP: P00740; LIXA.
 DR ZFIN: ZDB-GENE-990415-173; NOTCH.
 DR INTERPRO: IPR000152; .
 DR INTERPRO: IPR000561; .
 DR INTERPRO: IPR000800; .
 DR INTERPRO: IPR001336; .
 DR INTERPRO: IPR001438; .
 DR INTERPRO: IPR001881; .
 DR INTERPRO: IPR002110; .
 DR PFAM: PF00008; EGF_36.
 DR PFAM: PF00023; ank; 6.
 DR PFAM: PF00066; notch; 3.
 DR PRINTS: PR00009; EGFYGF.
 DR PRINTS: PR00010; EGFBLDOD.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS50297; ANK_RDP_REGION; 1.
 DR PROSITE: PS50010; ASA_HYDROXYL; 23.
 DR PROSITE: PS50022; EGF_1; 34.
 DR PROSITE: PS01186; EGF_2; 28.
 DR PROSITE: PS01187; EGF_CA; 22.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
 FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1725 1747 POTENTIAL.
 FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 57 EGF-LIKE 1.
 FT DOMAIN 58 98 EGF-LIKE 2.
 FT DOMAIN 101 138 EGF-LIKE 3.
 FT DOMAIN 139 175 EGF-LIKE 4.
 FT DOMAIN 177 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 371 409 EGF-LIKE 10.
 FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 489 524 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 526 562 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 564 599 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 601 637 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 639 674 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 676 712 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 714 749 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 751 787 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 827 865 EGF-LIKE 22.
 FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 981 1017 EGF-LIKE 26.
 FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1057 1093 EGF-LIKE 28.
 FT DOMAIN 1095 1141 EGF-LIKE 29.
 FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1265 1303 EGF-LIKE 33.
 FT DOMAIN 1305 1344 EGF-LIKE 34.
 FT DOMAIN 1346 1382 EGF-LIKE 35.
 FT DOMAIN 1385 1423 EGF-LIKE 36.
 FT DOMAIN 1446 1561 3 X LIN/NOTCH REPEATS.
 FT REPEAT 1446 1486 LIN/NOTCH 1.
 FT REPEAT 1487 1520 LIN/NOTCH 2.
 FT REPEAT 1521 1561 LIN/NOTCH 3.
 FT DOMAIN 1861 2074 6 X ANK MOTIF REPEATS.
 FT REPEAT 1861 1891 ANK MOTIF 1.
 FT REPEAT 1892 1940 ANK MOTIF 1.
 FT REPEAT 1941 1974 ANK MOTIF 1.
 FT REPEAT 1975 2007 ANK MOTIF 1.
 FT REPEAT 2008 2040 ANK MOTIF 1.
 FT REPEAT 2041 2074 ANK MOTIF 1.
 FT DOMAIN 2265 2276 POLY-GLN (OPA-REPEAT).
 FT DISULFID 25 35 BY SIMILARITY.
 FT DISULFID 29 45 BY SIMILARITY.
 FT DISULFID 47 56 BY SIMILARITY.
 FT DISULFID 62 73 BY SIMILARITY.
 FT DISULFID 67 86 BY SIMILARITY.
 FT DISULFID 88 97 BY SIMILARITY.
 FT DISULFID 105 116 BY SIMILARITY.
 FT DISULFID 110 126 BY SIMILARITY.
 FT DISULFID 128 137 BY SIMILARITY.
 FT DISULFID 143 154 BY SIMILARITY.
 FT DISULFID 148 163 BY SIMILARITY.
 FT DISULFID 165 174 BY SIMILARITY.
 FT DISULFID 181 194 BY SIMILARITY.
 FT DISULFID 188 203 BY SIMILARITY.
 FT DISULFID 205 214 BY SIMILARITY.
 FT DISULFID 221 232 BY SIMILARITY.
 FT DISULFID 226 242 BY SIMILARITY.
 FT DISULFID 244 253 BY SIMILARITY.
 FT DISULFID 260 271 BY SIMILARITY.
 FT DISULFID 265 280 BY SIMILARITY.
 FT DISULFID 282 291 BY SIMILARITY.
 FT DISULFID 298 311 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 338 349 BY SIMILARITY.
 FT DISULFID 343 358 BY SIMILARITY.
 FT DISULFID 360 369 BY SIMILARITY.
 FT DISULFID 375 386 BY SIMILARITY.
 FT DISULFID 380 397 BY SIMILARITY.
 FT DISULFID 399 408 BY SIMILARITY.
 FT DISULFID 415 428 BY SIMILARITY.
 FT DISULFID 422 437 BY SIMILARITY.
 FT DISULFID 439 448 BY SIMILARITY.
 FT DISULFID 455 466 BY SIMILARITY.
 FT DISULFID 460 475 BY SIMILARITY.
 FT DISULFID 477 486 BY SIMILARITY.
 FT DISULFID 493 503 BY SIMILARITY.
 FT DISULFID 498 512 BY SIMILARITY.
 FT DISULFID 514 523 BY SIMILARITY.
 FT DISULFID 530 541 BY SIMILARITY.
 FT DISULFID 535 550 BY SIMILARITY.
 FT DISULFID 552 561 BY SIMILARITY.
 FT DISULFID 568 578 BY SIMILARITY.
 FT DISULFID 573 587 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 605 616 BY SIMILARITY.
 FT DISULFID 610 625 BY SIMILARITY.
 FT DISULFID 627 636 BY SIMILARITY.
 FT DISULFID 643 653 BY SIMILARITY.
 FT DISULFID 648 662 BY SIMILARITY.
 FT DISULFID 664 673 BY SIMILARITY.
 FT DISULFID 680 691 BY SIMILARITY.
 FT DISULFID 695 700 BY SIMILARITY.
 FT DISULFID 702 711 BY SIMILARITY.
 FT DISULFID 718 728 BY SIMILARITY.

FT DISULFID 723 737 BY SIMILARITY.
 FT DISULFID 739 748 BY SIMILARITY.
 FT DISULFID 755 766 BY SIMILARITY.
 FT DISULFID 760 775 BY SIMILARITY.
 FT DISULFID 777 786 BY SIMILARITY.
 FT DISULFID 793 804 BY SIMILARITY.
 FT DISULFID 798 813 BY SIMILARITY.
 FT DISULFID 815 824 BY SIMILARITY.
 FT DISULFID 831 842 BY SIMILARITY.
 FT DISULFID 836 853 BY SIMILARITY.
 FT DISULFID 855 864 BY SIMILARITY.
 FT DISULFID 871 882 BY SIMILARITY.
 FT DISULFID 876 891 BY SIMILARITY.
 FT DISULFID 893 902 BY SIMILARITY.
 FT DISULFID 909 920 BY SIMILARITY.
 FT DISULFID 914 929 BY SIMILARITY.
 FT DISULFID 931 940 BY SIMILARITY.
 FT DISULFID 947 958 BY SIMILARITY.
 FT DISULFID 952 967 BY SIMILARITY.
 FT DISULFID 969 978 BY SIMILARITY.
 FT DISULFID 1023 1034 BY SIMILARITY.
 FT DISULFID 1038 1043 BY SIMILARITY.
 FT DISULFID 1045 1054 BY SIMILARITY.
 FT DISULFID 1061 1072 BY SIMILARITY.
 FT DISULFID 1066 1081 BY SIMILARITY.
 FT DISULFID 1083 1092 BY SIMILARITY.

Query Match 8.8%; Score 732.5; DB 1; Length 2437;
 Best Local Similarity 21.6%; Pred. No. 7.3e-38;
 Matches 323; Conservative 144; Mismatches 468; Indels 557; Gaps 66;

Qy 431 LAQNPFIDCHLW-----LADYLNPIETSGARCT----- 462
 Db 119 LUTDTPTCQCGSGKTCQADPCASNP-ANGGCSAFESHICTTCTPFGHQCTQD 177
 Qy 463 -----SPRLANKRIGQIKSKFR- -SOTEDYRSLSGDCPADLAPKRCBGT 512
 Db 178 VNECVASPSFCMGGTCTNEVGSYLCRPPETGPHQRCYQCL- -PSPCRSGGT 232
 Qy 513 VDCSN- -OKMLNPHIPIQYLAELNNEPTVLEATGIFPKLPQLRKSNKNTDIE 570
 Db 233 VQTSOTHTCSCLPGFTQG- -TCENVDDCTQACENG- -PCIDGINTYHCDKXN 286
 Qy 571 BGAFCASGVNEILLTSNLEN- -VQRMFKGLSEI- - - - - 604
 Db 287 TGOI-CTEDVDEELSPNACQNGTCHTIGGKFCVNGWTDGSCNIDCASAACH 345
 Qy 605 -----KTLMLRSNRITCVNGDSFGLSSVLLSYDNQITTVAGAF 646
 Db 346 GATCHDRVASFPCBCPHGTGLLHLDACISNPGCK- -SNCTNPVSGKAICTPPGT 404
 Qy 647 DTL- -HSLSTLNLNLP- - - - -FNCNYLAWGEMLKKRIRIVTGNPQKPY 691
 Db 405 GSACNDIDECISLANPCEHGRCLNTKGSFQCLQGYG- -PREM- - - - - 450
 Qy 692 FLAEIPIQVADQPTCDGDDNSCSPLSRCTECTCLDTV- -VRCS- 737
 Db 451 -----DVNECKS- -NPCQNDATCLDQIGGFCICMPGTBGFVFCI 488
 Qy 738 -----NGLKVLVPGKIPROVTLYLD- - - - -GNQFTLVPEL 769
 Db 489 NSDQSCQPLNGKCIDKINSFHCPCPGSGSLQVQVDECASTPCNKGACTDGP- - 545
 Qy 770 SNTXHLTLIDLSNNRISTLSNQSPS- - - - -NMQLTLILSYNRLACIPPTFDGLASL 823
 Db 546 -----NKYTCTCTPGFSGEICEDLINECASSPCHYGVCR- - - - -DGVASF 585
 Qy 824 -----RLLSHGNDISVVP- - - - -EGAFVLDLSALHAIQNPVLCQNM 863
 Db 586 TDCRPGVGTGRCLNTWINECLSQPCNNGSTQDRENN- - - - -ITCTPSTTGVNKEINI 640
 Qy 864 QWLS- - - - -DVKSEYK- - - - -EPG- - - - -IARCAQGVDMAKLLTTP 897
 Db 897 -----

Db 641 DCKRKPCTQYKCIDKINGCYCEPVSFGSMCNINIDDCALNPNCHNGSTCIDGV----- 695
 Qy 898 SKAFTC-----QGPVDVNLAKNCPCLSNPNCKNDSPDVPDYRCTCTPGFGQCCDVP1 954
 Db 696 -NSTPCLCPDGFDRATCLSQHNECSNPCH-GSC-LQJINSYRVCAGWGRKIDINI 752
 Qy 955 HACSINPCNHGCTCHLAGEEDGFWICADGFECEVNWDDCNDENNSTCVGIN 1014
 Db 753 NECLSNPCVNGTCT- - -KMTSGYLCTCRAGFSQPMQCNINECASNPCLNQSGCIDVA 809
 Qy 1015 NYTCLCPPTTGELCE- - - - -ERLDF- - - - -CAQDL- 1040
 Db 810 GFYCNMLPTTGEVCEVNLAPCSPPRCXNGVYCRESDPQSFSCNPGWQGTCEVUN 869
 Qy 1041 -----NPODCSKCLTPKGFKCDCTPGYVGEHCDIFDDQDN- - - - - 1079
 Db 870 BCVWNPCTGVNGVCENLGGPQCRMGPTGALCENIDDCENPNCSSNGVQDRVWFCV 929
 Qy 1080 -----KCKNAGACTDANVGTCTCPGVSGLCEFSPPFWLPRT 1118
 Db 930 VCLAGPFGKCAEDIDCVSAPCRNGMCTDCVNSTCSCPAGFGSGCENIWP- - - - - 983
 Qy 1119 SPQCNQDQWQAQIVRIINEPICQLPGYQSGCEKXLSVWFIWKSQYIIPSAVRPQT 1178
 Db 984 -DCTESSCPWGTGCVDSIGSFSCVCLPGFTGNTQGS- - - - -DVNECDSPQ 1029
 Qy 1179 NITLIATDEDSGLLYKGDHIAVEL-YGRVY- - - - -A 1213
 Db 1030 N-----GGSCDGTGTCTCTPGTGTGLNCSYVWDCSSPKNGGSCWQAGSTQCA 1084
 Qy 1214 SYDTGSPASAIT-SVETINDGNFHVIELLALQSLSL- - - - -VDGNGFIITNL 1263
 Db 1085 SGWTG- - - - -TYCDVFSV- - - - -CVAARQGVSAVLCHRAGQCVDAWG- - - - -TFL 1128
 Qy 1264 SKQSTLNFOSPLTVGGMPG- - - - -KSNVSLRAQPGNGTS- - - - -FNGCIR 1305
 Db 1129 CRQ- - - - -AGTGTSCQGVDCQPNQCGATCTDYLGGTSCEVPGTHG- - - - - 1175
 Qy 1306 NLYINSELQDFQVPMQGTILPGCEPC- - - - -HKVCAHGT- - - - -CQPSQ- 1347
 Db 1176 -MCKSEINECLSPQCG- - - - -GTCLDINVTYKSCPGTQGVHCEIDIDCSPVDP 1229
 Qy 1348 -----AGTCTCBQBMWGLCDQRTNPLGLMKVHGTCPLINAFS- - - - - 1388
 Db 1230 LTGEPCTFNGRCVDRVGGYGVCPAGVGRCEGDVNE- - - - -CLSDPCDPGSGTNCV 1283
 Qy 1389 ----YSCKLBGGHVLCEDEEDLNPQAIKCKH-KGKRL-SGLGQY-CBSCSGTG 1440
 Db 1284 QLINDRCPCRTGTGKRC- - - - -ETVYCNCKTPTCKNGQCAVASNTHGTCIKCPGTS 1340
 Qy 1441 DSCREI- -SCHGERIRDYKQKGYAA- - - - -CQTKYKSLER- - - - -GGCAGQCC 1488
 Db 1341 SSCETQSSQSLACRNGATCVSGLSRPLCAFGPSGECQTRNDSPCLVNPTCVNGT 1400
 Qy 1489 GPLRSKR- - - - -RKYSECTDGSFVDEKVKWKCCTRC 1523
 Db 1401 QPISDAPPTKCSFANFNGLCHILDYFSFGQGRDIAPVVEIRCEIAQC 1452

RESULT 8
 FBPLSTRPU
 ID FBPLSTRPU STANDARD; PRT: 1064 AA.
 AC P10079;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-FEB-1995 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE FIBROPELIN 1 PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
 DE (DGF-1).
 GN EGF1.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Echinodermata; Echinodermata; Eleutherozoa; Echinozoa;
 OC Chordata; Echinodermata; Echinacea; Echinoidae; Strongylocentrotidae;
 OC Strongylocentrotus.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE+90112459; PubMed-2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT 'Structural analysis of the uEGF gene in the sea urchin
RT stronglyocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.';
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE-87319677; PubMed-3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT 'A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.';
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE+89196806; PubMed-2784773;
RA Hunt L.T., Barker W.C.;
RT 'Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin.';
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE+91285254; PubMed-2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT 'Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.';
RL Dev. Biol. 146:89-99(1991).
CC -I- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: IA (SHOWN HERE) AND IB; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF
CC REPEATS.
CC -I- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -I- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -I- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08692; AAA62164.1; .
DR EMBL: L08692; AAA62163.1; .
DR EMBL: X17530; CAA35571.1; .
DR EMBL: M17421; AAA30050.1; .
DR EMBL: X17533; CAA35573.1; .
DR PIR: A29316; A29316.
DR HSP: P01132; IEPH.
DR INTERPRO: IPR000088; .
DR INTERPRO: IPR000152; .
DR INTERPRO: IPR000561; .
DR INTERPRO: IPR000859; .
DR INTERPRO: IPR001438; .
DR INTERPRO: IPR001881; .
DR PFAM: PF01382; Avidin; 1.
DR PFAM: PF00431; CUB; 1.
DR PFAM: PF00008; EGF; 21.

DR PRINTS; PRO0010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; 19.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 19.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
KW Glycoprotein.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1064 FIBROPELLEIN I.
FT DOMAIN 20 55 EGF-LIKE 1.
FT DOMAIN 62 175 CUB.
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 860 896 EGF-LIKE 20.
FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 936 1064 AVIDIN-LIKE.
FT DISULFID 23 34 BY SIMILARITY.
FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 240 249 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 287 BY SIMILARITY.
FT DISULFID 294 305 BY SIMILARITY.
FT DISULFID 299 314 BY SIMILARITY.
FT DISULFID 316 325 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.

Ov 1408 -----LFNPCOAICKKHGKRLSGLGOPYCECSSGYTGDSCDREI-SCRG----- 1451

```

CC      FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATORIAL ROLE DURING
CC      VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC      DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC      -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC      -2- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC      -3- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC      -4- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC      -5- SIMILARITY: CONTAINS 6 CND10/SWI6 REPEATS.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: X74760; CAA52776.1; -.
DR      HSSP: P00740; LIXA.
DR      MGD: MGI:99460; NOTCH3.
DR      INTERPRO: IPR000152; -.
DR      INTERPRO: IPR000561; -.
DR      INTERPRO: IPR000800; -.
DR      INTERPRO: IPR001438; -.
DR      INTERPRO: IPR001881; -.
DR      INTERPRO: IPR002110; -.
DR      PFAM: PF00008; EGF; 34.
DR      PFAM: PF00023; ank; 6.
DR      PFAM: PF00066; notch3.
DR      PRINTS: PR00010; GCFBL00.
DR      PROSITE: PS50088; ANK_REPEAT; 4.
DR      PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR      PROSITE: PS00010; ASX_HYDROXYL; 18.
DR      PROSITE: PS00022; EGF_L1; 13.
DR      PROSITE: PS00186; EGF_2; 17.
DR      PROSITE: PS00187; EGF_CA; 12.
KW      Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW      Glycoprotein.
FT      DOMAIN      1      1643      EXTRACELLULAR.
FT      TRANSMEM    1644      1654      POTENTIAL.

```

FT DOMAIN	1665	2318	CYTOPLASMIC.
FT DOMAIN	39	1374	34 X EGF-TYPE REPEATS.
FT DOMAIN	1388	1503	3 X LIN/NOTCH REPEATS.
FT DOMAIN	1784	1998	6 X CDC10/SW16 REPEATS.
FT DOMAIN	2242	2261	PEST.
FT DOMAIN	39	78	EGF-LIKE 1.
FT DOMAIN	79	119	EGF-LIKE 2.
FT DOMAIN	120	157	EGF-LIKE 3.
FT DOMAIN	159	196	EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	198	235	EGF-LIKE 5.
FT DOMAIN	237	273	EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	275	313	EGF-LIKE 7.
FT DOMAIN	315	351	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	352	390	EGF-LIKE 9.
FT DOMAIN	392	430	EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	432	468	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	470	506	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	508	544	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	546	581	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	583	619	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	621	656	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	658	694	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	696	731	EGF-LIKE 18.
FT DOMAIN	735	771	EGF-LIKE 19.
FT DOMAIN	772	809	EGF-LIKE 20.
FT DOMAIN	811	848	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	850	886	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	888	923	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	925	961	EGF-LIKE 24.
FT DOMAIN	963	999	EGF-LIKE 25.
FT DOMAIN	1001	1035	EGF-LIKE 26.
FT DOMAIN	1037	1083	EGF-LIKE 27.
FT DOMAIN	1085	1121	EGF-LIKE 28.
FT DOMAIN	1123	1159	EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1161	1204	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1206	1245	EGF-LIKE 31.
FT DOMAIN	1247	1288	EGF-LIKE 32.
FT DOMAIN	1290	1326	EGF-LIKE 33.
FT DOMAIN	1336	1374	EGF-LIKE 34.
FT REPEAT	1388	1428	LIN/NOTCH 1.
FT REPEAT	1429	1467	LIN/NOTCH 2.
FT REPEAT	1468	1503	LIN/NOTCH 3.
FT REPEAT	1784	1816	CDC10/SW16 1.
FT REPEAT	1817	1865	CDC10/SW16 2.
FT REPEAT	1866	1898	CDC10/SW16 3.
FT REPEAT	1899	1932	CDC10/SW16 4.
FT REPEAT	1933	1965	CDC10/SW16 5.
FT REPEAT	1966	1998	CDC10/SW16 6.
FT DISULFID	43	55	BY SIMILARITY.
FT DISULFID	49	66	BY SIMILARITY.
FT DISULFID	68	77	BY SIMILARITY.
FT DISULFID	83	94	BY SIMILARITY.
FT DISULFID	88	107	BY SIMILARITY.
FT DISULFID	109	118	BY SIMILARITY.
FT DISULFID	124	135	BY SIMILARITY.
FT DISULFID	129	145	BY SIMILARITY.
FT DISULFID	147	156	BY SIMILARITY.
FT DISULFID	163	175	BY SIMILARITY.
FT DISULFID	169	184	BY SIMILARITY.
FT DISULFID	186	195	BY SIMILARITY.
FT DISULFID	202	213	BY SIMILARITY.
FT DISULFID	207	223	BY SIMILARITY.
FT DISULFID	225	234	BY SIMILARITY.
FT DISULFID	241	252	BY SIMILARITY.
FT DISULFID	246	261	BY SIMILARITY.
FT DISULFID	263	272	BY SIMILARITY.
FT DISULFID	279	292	BY SIMILARITY.
FT DISULFID	286	301	BY SIMILARITY.
FT DISULFID	303	312	BY SIMILARITY.
FT DISULFID	319	330	BY SIMILARITY.
FT DISULFID	324	339	BY SIMILARITY.
FT DISULFID	341	350	BY SIMILARITY.
FT DISULFID	356	367	BY SIMILARITY.

FT DISULFID	361	378	BY SIMILARITY.
FT DISULFID	380	389	BY SIMILARITY.
FT DISULFID	396	409	BY SIMILARITY.
FT DISULFID	403	418	BY SIMILARITY.
FT DISULFID	420	429	BY SIMILARITY.
FT DISULFID	436	447	BY SIMILARITY.
FT DISULFID	441	456	BY SIMILARITY.
FT DISULFID	458	467	BY SIMILARITY.
FT DISULFID	474	485	BY SIMILARITY.
FT DISULFID	479	494	BY SIMILARITY.
FT DISULFID	496	505	BY SIMILARITY.
FT DISULFID	512	523	BY SIMILARITY.
FT DISULFID	517	532	BY SIMILARITY.
FT DISULFID	534	543	BY SIMILARITY.
FT DISULFID	550	560	BY SIMILARITY.
FT DISULFID	555	569	BY SIMILARITY.
FT DISULFID	571	580	BY SIMILARITY.
FT DISULFID	587	598	BY SIMILARITY.
FT DISULFID	592	607	BY SIMILARITY.
FT DISULFID	609	618	BY SIMILARITY.
FT DISULFID	625	635	BY SIMILARITY.
FT DISULFID	630	644	BY SIMILARITY.
FT DISULFID	646	655	BY SIMILARITY.
FT DISULFID	662	673	BY SIMILARITY.
FT DISULFID	667	682	BY SIMILARITY.
FT DISULFID	684	693	BY SIMILARITY.
FT DISULFID	700	710	BY SIMILARITY.
FT DISULFID	705	719	BY SIMILARITY.
FT DISULFID	721	730	BY SIMILARITY.
FT DISULFID	739	750	BY SIMILARITY.
FT DISULFID	744	759	BY SIMILARITY.
FT DISULFID	761	770	BY SIMILARITY.
FT DISULFID	776	787	BY SIMILARITY.
FT DISULFID	781	797	BY SIMILARITY.
FT DISULFID	799	808	BY SIMILARITY.
FT DISULFID	815	827	BY SIMILARITY.
FT DISULFID	821	836	BY SIMILARITY.
FT DISULFID	838	847	BY SIMILARITY.
FT DISULFID	854	865	BY SIMILARITY.
FT DISULFID	859	874	BY SIMILARITY.
FT DISULFID	876	885	BY SIMILARITY.
FT DISULFID	892	902	BY SIMILARITY.
FT DISULFID	897	911	BY SIMILARITY.
FT DISULFID	913	922	BY SIMILARITY.
FT DISULFID	929	940	BY SIMILARITY.
FT DISULFID	934	949	BY SIMILARITY.
FT DISULFID	951	960	BY SIMILARITY.
FT DISULFID	967	978	BY SIMILARITY.
FT DISULFID	972	987	BY SIMILARITY.
FT DISULFID	989	998	BY SIMILARITY.
FT DISULFID	1005	1016	BY SIMILARITY.
FT DISULFID	1010	1023	BY SIMILARITY.
FT DISULFID	1025	1034	BY SIMILARITY.
FT DISULFID	1041	1062	BY SIMILARITY.
FT DISULFID	1056	1071	BY SIMILARITY.
FT DISULFID	1073	1082	BY SIMILARITY.
FT DISULFID	1089	1100	BY SIMILARITY.
FT DISULFID	1094	1109	BY SIMILARITY.
FT DISULFID	1111	1120	BY SIMILARITY.
FT DISULFID	1127	1138	BY SIMILARITY.
FT DISULFID	1132	1147	BY SIMILARITY.
FT DISULFID	1149	1158	BY SIMILARITY.
FT DISULFID	1165	1183	BY SIMILARITY.
FT DISULFID	1177	1192	BY SIMILARITY.
FT DISULFID	1194	1203	BY SIMILARITY.
FT DISULFID	1210	1223	BY SIMILARITY.
FT DISULFID	1215	1233	BY SIMILARITY.
FT DISULFID	1235	1244	BY SIMILARITY.

Query Match 8.7%; Score 722; DB 1; Length 2318;
 Best Local Similarity 23.7%; Pred. No. 3.1e-37;
 Matches 227; Conservative 86; Mismatches 288; Indels 358; Gaps 37;

GN PROTEIN INT-3.
 OS NOTCH4 OR INT3 OR INT-3.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gellahan D., Callahan R.;
 RT 'Mouse mammary tumor gene int-3: a member of the notch gene family
 RT transforms mammary epithelial cells.';
 RL J. Virol. 66:2594-2599(1992).
 RN [2]
 RP REVISIONS. SEQUENCE FROM N.A.
 RX MEDLINE=97294599; PubMed=9150355;
 RA Gellahan D., Callahan R.;
 RT 'The mouse mammary tumor associated gene INT3 is a unique member of
 RT the NOTCH gene family (NOTCH4).';
 RL Oncogene 14:1883-1890(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LONG, AND TESTIS;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uyttendaele S., Maraziti G., Wu G., Yan Q., Sassoon D., Ritsajewski J.;
 RT 'Notch4/Int-3, a mammary proto-oncogene, is an endothelial
 RT cell-specific mammalian Notch gene.';
 RL Development 122:2251-2259(1996).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIM/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 CDC10/SWIS REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENCL: M80456; AAB38377.1; -.
 DR ENCL: U43691; ACS2630.1; -.
 DR PIR: A38072; TWM73.
 DR HSP: P00740; 1X1.
 DR MGD: MGI:107471; NOTCH4.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR000800; -.
 DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001881; -.
 DR INTERPRO: IPR002110; -.
 DR PFAM: PF00008; EGF; 27.
 DR PFAM: PF00023; ank; 6.
 DR PFAM: PF00066; notch; 2.
 DR PRINTS: PR00010; EGFBLD.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 11.
 DR PROSITE: PS00022; EGF_1; 28.
 DR PROSITE: PS01186; EGF_2; 21.
 DR PROSITE: PS01187; EGF_CA; 9.

```

RESULT 10
NTC4_MOUSE
ID NTC4_MOUSE STANDARD; PRT; 1964 AA.
AC P31695; Q62389;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING

```

FT	DISU01FD	514	525	BY	SIMILARITY
FT	DISU01FD	519	534	BY	SIMILARITY
FT	DISU01FD	536	545	BY	SIMILARITY
FT	DISU01FD	552	563	BY	SIMILARITY
FT	DISU01FD	557	572	BY	SIMILARITY
FT	DISU01FD	574	583	BY	SIMILARITY
FT	DISU01FD	590	601	BY	SIMILARITY
FT	DISU01FD	595	610	BY	SIMILARITY
FT	DISU01FD	612	621	BY	SIMILARITY
FT	DISU01FD	626	637	BY	SIMILARITY
FT	DISU01FD	631	646	BY	SIMILARITY
FT	DISU01FD	648	655	BY	SIMILARITY
FT	DISU01FD	662	669	BY	SIMILARITY
FT	DISU01FD	664	674	BY	SIMILARITY
FT	DISU01FD	676	685	BY	SIMILARITY
FT	DISU01FD	692	703	BY	SIMILARITY
FT	DISU01FD	697	712	BY	SIMILARITY
FT	DISU01FD	714	723	BY	SIMILARITY
FT	DISU01FD	730	741	BY	SIMILARITY
FT	DISU01FD	735	750	BY	SIMILARITY
FT	DISU01FD	752	761	BY	SIMILARITY
FT	DISU01FD	768	779	BY	SIMILARITY
FT	DISU01FD	773	788	BY	SIMILARITY
FT	DISU01FD	790	799	BY	SIMILARITY
FT	DISU01FD	807	818	BY	SIMILARITY
FT	DISU01FD	812	827	BY	SIMILARITY
FT	DISU01FD	829	838	BY	SIMILARITY
FT	DISU01FD	845	856	BY	SIMILARITY
FT	DISU01FD	850	865	BY	SIMILARITY
FT	DISU01FD	867	876	BY	SIMILARITY
FT	DISU01FD	882	903	BY	SIMILARITY
FT	DISU01FD	897	912	BY	SIMILARITY
FT	DISU01FD	914	923	BY	SIMILARITY
FT	DISU01FD	930	941	BY	SIMILARITY
FT	DISU01FD	935	950	BY	SIMILARITY
FT	DISU01FD	952	961	BY	SIMILARITY
FT	DISU01FD	968	979	BY	SIMILARITY
FT	DISU01FD	973	988	BY	SIMILARITY
FT	DISU01FD	990	999	BY	SIMILARITY
FT	DISU01FD	1006	1019	BY	SIMILARITY
FT	DISU01FD	1011	1028	BY	SIMILARITY
FT	DISU01FD	1030	1039	BY	SIMILARITY
FT	DISU01FD	1046	1057	BY	SIMILARITY
FT	DISU01FD	1051	1069	BY	SIMILARITY
FT	DISU01FD	1071	1080	BY	SIMILARITY
FT	DISU01FD	1087	1098	BY	SIMILARITY
FT	DISU01FD	1092	1110	BY	SIMILARITY
FT	DISU01FD	1112	1121	BY	SIMILARITY

Query Match . 8.3%; Score 691.5; DB 1; Length 1964;
Best Local Similarity 24.8%; Pred. No. 2e-35;
Matches 235; Conservative 86; Mismatches 345; Indels 283; Gaps 41;

QY	664	CNYLNLNLGEHLKRAIKAVITGNPRCKPYFLKEIPQDAVADPTCDGDNDSNC-----	718
Db	48	CQCAQSPFLGE-----TCQFP-----DPCRDQTL-----CKNG--GSQALLPT	83
QY	719	-PLSRCPF-----BCTCLDVTVRCSNKGKLVLPKGIPTDWITLLDNGQFTVPKELS	770
Db	64	PPSSSPSPFLPHFSCT-----PSGFTGDRCTHLEE--LCPPSFC	125
QY	771	NYKHLTLDLSNNRISLTLSQNSQNWQL-----TLLISYNLRICPPHPTD	818
Db	126	NGGH-CIVAGSRPQSCPEGWGTCQQLDRFCSNACNANGVCLATIPQICRCPPGE	184
QY	819	GLKSLRLSLSGNDISVYPEGAFNDLSALSHAIATGNVYCDCCNWLSDWVKSEYKPE	878
Db	185	GHTCEDRINCFLPECPQGS-----PSCNLTLSGYQCPLVGGQGPQCKLRGACGCP	238
QY	879	IARACG-----PGEMAKLLITPSKFKTCGGPVDVNLAKNC--CLSNPKNDGTCN	930
Db	239	SLCNGTGQCLVPGHSTHFLCLCPG--FT-----GLDCENPDCTVRCPOGATC-	288

```

RESULT 11
CRB_DROME
ID CRB_DROME STANDARD; PRT: 2139 AA.
AC P10040;
RT 01-MAR-1989 (Rel. 10, Created)
RT 01-MAY-1991 (Rel. 18, Last sequence update)
RT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CRUMBS PROTEIN PRECURSOR (95F).
GN CRB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RT SEQUENCE FROM N.A.
RC STRAIN-OREGON-R; TISSUE-EMBRYO;
RX MEDLINE=90263104; PubMed=234615;
RA Tepass U., Theres C., Knust E.
RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT Drosophila epithelial cells and required for organization of
RT epithelia."
RL Cell 61:787-799(1990).
RN [2]
RT SEQUENCE OF 1663-1955 FROM N.A.
RX MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
RA Vassin H., Campos-Ortega J.A.;
RT "RGF homologous sequences encoded in the genome of Drosophila

```

melanogaster, and their relation to neurogenic genes.*;
 RL EMBL J. 6:761-766(1987).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
 CC POLARITY. IT MAY ACT AS A SIGNAL.
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.
 CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: M33753; AAA28428.1; ALT_SEQ.
 DR EMBL: X05144; CAA28793.1; -.
 DR PIR: B26637; B26637.
 DR PIR: A35672; A35672.
 DR HSP: P00740; 1TAA.
 DR FLIBASE: Fbg0000368; crb.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001791; -.
 DR INTERPRO: IPR001881; -.
 DR PFAM: PF00008; EGF; 27.
 DR PFAM: PF00054; laminin_G; 3.
 DR PRINTS: PR00010; EGFBL00D.
 DR PROSITE: PS00010; ASK_HYDROXYL; 15.
 DR PROSITE: PS00022; EGF_L_1; 26.
 DR PROSITE: PS01186; EGF_2; 17.
 DR PROSITE: PS01187; BGF_C; 15.
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 90
 FT CHAIN 1 2139 CRUMBS PROTEIN.
 FT DOMAIN 1 2084 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2085 2111 POTENTIAL.
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 267 303 EGF-LIKE 1.
 FT DOMAIN 306 343 EGF-LIKE 2.
 FT DOMAIN 348 386 EGF-LIKE 3.
 FT DOMAIN 388 425 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 427 463 EGF-LIKE 5.
 FT DOMAIN 464 500 EGF-LIKE 6.
 FT DOMAIN 501 532 EGF-LIKE 7.
 FT DOMAIN 545 581 EGF-LIKE 8.
 FT DOMAIN 582 611 EGF-LIKE 9.
 FT DOMAIN 609 646 EGF-LIKE 10.
 FT DOMAIN 648 685 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 687 723 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 725 761 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 763 800 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 802 838 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 840 902 EGF-LIKE 16.
 FT DOMAIN 904 940 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 942 978 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 980 1021 EGF-LIKE 19.
 FT DOMAIN 1207 1243 EGF-LIKE 20.
 FT DOMAIN 1481 1517 EGF-LIKE 21.
 FT DOMAIN 1759 1795 EGF-LIKE 22.
 FT DOMAIN 1797 1833 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1835 1871 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1874 1915 EGF-LIKE 25.
 FT DOMAIN 1915 1951 EGF-LIKE 26.
 FT DOMAIN 1953 1989 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1991 2029 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 2030 2070 EGF-LIKE 29.
 FT DISULFID 271 282 BY SIMILARITY.
 FT DISULFID 276 291 BY SIMILARITY.

FT	DISULFID	1919	1930	BY SIMILARITY.
FT	DISULFID	1924	1939	BY SIMILARITY.
FT	DISULFID	1941	1950	BY SIMILARITY.
FT	DISULFID	1957	1968	BY SIMILARITY.
FT	DISULFID	1962	1977	BY SIMILARITY.
FT	DISULFID	1979	1988	BY SIMILARITY.
FT	DISULFID	1995	2008	BY SIMILARITY.
FT	DISULFID	2002	2017	BY SIMILARITY.
FT	DISULFID	2019	2028	BY SIMILARITY.
FT	CARBOHYD	37	37	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	96	96	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	238	238	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	336	336	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	400	400	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	550	550	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	565	565	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	736	736	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	746	746	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	860	860	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	884	884	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	976	976	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	1102	1102	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	1114	1114	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	1138	1138	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	1192	1192	N-LINKED (GLCNAC. .) (POTENTIAL)
FT	CARBOHYD	1245	1245	N-LINKED (GLCNAC. .) (POTENTIAL)

Query Match 7.6%; Score 634.5; DB 1; Length 2139;
Best Local Similarity 24.3%; Pred. No. 8.4e-32;
Matches 226; Conservative 91; Mismatches 273; Indels 341; Gaps 45

[illegible]

Qy 1121 C-----DNFDCQ-----NGAQC1-VRINEP-----IQCC 1143
 Db 952 CQRGCTCQDIDYDCDCDAMYGKNCVLLKGCQDQCLNGAGACLPYLINVEVHTCTC 1011
 Qy 1144 LQGVQCKEKK-----LYSVNFINKSEYLIQPSAIVKQPTNITLQIATEDSGIL 1193
 Db 1012 ENGFGQDKCKERTTILSNWATSLSVITTEBEGY-----DINLQFTTILPGLVL 1059
 Qy 1194 LY--KGDKH-----JAVELYGRVRSYDTGSHPSAIVSVETINDGNFHVELLQSL 1248
 Db 1060 AFGTTGSEKNEPVSYLLELJLRLNL-HSSLNKGVEFVIGSKLNQSNHXY-FAVNTS- 1116
 Qy 1249 SLSVDGNGKILITLKSQSLNFPSP-----LYVGG-MPG-KSNVASLQAQPGNGTSPHG 1302
 Db 1117 HLVLASNDQAIPPVGSEYETANNQSPSPRYLGGTIPMLKSYLRLHLPQ-----SAFVG 1172
 Qy 1303 CIRMLINSE--LQDPQVPMQ-----TGILPGC-----EPCHKV-CA---HGTCPQ 1344
 Db 1173 CMOQILWGWKGVFPEDQANISYTKLENVQSGCPRTBQCKWPCNSWECTDLNH----- 1227
 Qy 1345 SSQAGFTCEGCGEWMPLCDQRTNDPCIGNK 1375
 Db 1228 -----TFACPCRPFPFGTCHQNMATPQHE 1254

RESULT 12
 SERR_DROME
 ID: SERR_DROME STANDARD; PRT: 1408 AA.
 AC P18168;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SERRATE PROTEIN PRECURSOR (BEADED PROTEIN).
 GN SERR OR BD.
 OS Drosophila melanogaster (Fruit fly).
 CC Bakaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORDGON-R;
 RX MEDLINE=91347903; PubMed=1840519;
 RA Thomas U., Speicher S.A., Knust E.;
 RT "The Drosophila gene Serrate encodes an EGF-like transmembrane
 RT protein with a complex expression pattern in embryos and wing
 RT discs";
 RL Development 111:749-761(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91099666; PubMed=2125287;
 RA Fleming R.J., Scottgale T.N., Diederich R.J., Artavanis-Tsakonas S.;
 RT "The gene Serrate encodes a putative EGF-like transmembrane protein
 RT essential for proper ectodermal development in Drosophila
 RT melanogaster";
 RL Genes Dev. 4:2188-2201(1990).
 CC -1- FUNCTION: ESSENTIAL FOR PROPER ECTODERMAL DEVELOPMENT. SERRATE
 CC MAY REPRESENT AN ELEMENT IN A NETWORK OF INTERACTING MOLECULES
 CC OPERATING AT THE CELL SURFACE DURING THE DIFFERENTIATION OF
 CC CERTAIN TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: APPEARS TO BE RESTRICTED EXCLUSIVELY TO
 CC CELLS OF ECTODERMAL ORIGIN.
 CC -1- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.
 CC -1- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN
 CC LEVEL, IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY
 CC COMPETE FOR BINDING WITH THE NOTCH PROTEIN.
 CC -1- SIMILARITY: CONTAINS 14 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE DELTA/SERRATE/JAGGED FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X56811; CAA00148.1;
 DR EMBL: M35759; AAA28938.1;
 DR PIR: A36666; A36666.
 DR PIR: S16878; S16878.
 DR HSSP: P00743; LWEE.
 DR FLTBASE: F8g0004197; Ser.
 DR INTERPRO: IPR000152;
 DR INTERPRO: IPR000561;
 DR INTERPRO: IPR001438;
 DR INTERPRO: IPR001774;
 DR INTERPRO: IPR001881;
 DR PFAM: PF01414; BSL; 1.
 DR PFAM: PF00008; EGF; 11.
 DR PRINTS: PR00010; EGFLEOOD.
 DR PROSITE: PS00010; ASX_HYDROXYL; 7.
 DR PROSITE: PS00022; EGF_1; 14.
 DR PROSITE: PS01186; EGF_2; 8.
 DR PROSITE: PS01187; EGF_CA; 5.
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 83 POTENTIAL.
 FT CHAIN 84 1408 SERRATE PROTEIN.
 FT DOMAIN 84 1223 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1224 1249 POTENTIAL.
 FT DOMAIN 1250 1408 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 284 317 EGF-LIKE 1.
 FT DOMAIN 315 349 EGF-LIKE 2.
 FT DOMAIN 351 389 EGF-LIKE 3.
 FT DOMAIN 391 489 EGF-LIKE 4.
 FT DOMAIN 407 476 SER-RICH (INSERT).
 FT DOMAIN 491 527 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 529 609 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 611 646 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 648 684 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 686 721 EGF-LIKE 9.
 FT DOMAIN 723 797 EGF-LIKE 10.
 FT DOMAIN 737 769 THR-RICH (INSERT).
 FT DOMAIN 799 835 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 837 877 EGF-LIKE 12.
 FT DOMAIN 879 915 EGF-LIKE 13.
 FT DOMAIN 917 953 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 288 299 BY SIMILARITY.
 FT DISULFID 292 305 BY SIMILARITY.
 FT DISULFID 307 316 BY SIMILARITY.
 FT DISULFID 319 330 BY SIMILARITY.
 FT DISULFID 325 337 BY SIMILARITY.
 FT DISULFID 339 348 BY SIMILARITY.
 FT DISULFID 355 367 BY SIMILARITY.
 FT DISULFID 361 377 BY SIMILARITY.
 FT DISULFID 379 388 BY SIMILARITY.
 FT DISULFID 395 406 BY SIMILARITY.
 FT DISULFID 400 477 BY SIMILARITY.
 FT DISULFID 479 488 BY SIMILARITY.
 FT DISULFID 495 506 BY SIMILARITY.
 FT DISULFID 500 515 BY SIMILARITY.
 FT DISULFID 517 526 BY SIMILARITY.
 FT DISULFID 533 588 BY SIMILARITY.
 FT DISULFID 582 597 BY SIMILARITY.
 FT DISULFID 599 608 BY SIMILARITY.
 FT DISULFID 615 625 BY SIMILARITY.
 FT DISULFID 619 634 BY SIMILARITY.
 FT DISULFID 636 645 BY SIMILARITY.
 FT DISULFID 652 663 BY SIMILARITY.
 FT DISULFID 657 672 BY SIMILARITY.
 FT DISULFID 674 683 BY SIMILARITY.
 FT DISULFID 690 700 BY SIMILARITY.

```

RESULT 13
L112.CAEEL
ID L112.CAEEL STANDARD: PRT: 1429 AA.
AC P14585;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LIN-12 PROTEIN PRECURSOR.
GN LIN-12 OR R107.3
OS Caenorhabditis elegans.
OC Bukaryota; Metazoa; Nematoa; Chromodorea; Rhabditia; Rhabditoidea;
OC Rhabditidae; Pelodineae; Caenorhabditis.
RN
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=88334747; PubMed=3419531;
RY Yochim J., Weston K., Greenwald I.;
RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
RP protein with overall similarity to Drosophila Notch.";
RL Nature 335:547-550(1988).
RN
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnoughosh R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copestey T., Cooper J., Coulson A.
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Perry C., Rifken L., Roopra A., Saunders D., Showknee R.
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thery-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wolhaden P.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-33(1994).
CC
CC -1- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT
CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.
CC
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC
CC -1- SIMILARITY: HIGH. TO C.ELEGANS GLP-1.
CC
CC -1- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
CC
CC -1- SIMILARITY: CONTAINS 5 LIN/NOTCH REPEATS.
CC
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABO-
CC RATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COM-
CC MUNITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/ann
CC OR SEND AN EMAIL TO license@isb-sib.ch).
CC

```

```

QY 648 LHLSTLNLN-----ANFP-----CKYLAWLGLWAKKRIYVGNPQKPYE 692
Db 18 SLRIQSGLLGLGCRNGCHAGVNPQVTSYWCRCDBGFGGYSQCCQDVKSGADKCEVF 77

QY 693 LKRIPIQVVALQDPTDC-----DGDNDNSCSPLSRPHECTLDITVRCNS 738
Db 78 DRYRYMETCVCKD--CDINGNSLLKPSCPSGYGGD-----CKYQGTQPSV--CMN 125

QY 739 GLKLVAKGIPRVDYELLDGQFTLVPEKLSYKHLTLDLSNNRISLWSQSFNNMQ 798
Db 126 GG----- 127

QY 799 LLTLLTSNLRACIPPTDGLKALLSLHNGNDSVYVPEGAFNOLSLAHLAGNPLY 858
Db 128 ---CQTGAAGNACAC---DGFKEG-CELDVHECEENKACNGRSTCWTL---GTVI 177

QY 859 CDCNMQWLSDDWVKSETEPGIARCAKCGEMADKLLLTSPSKFTQCGPVVNLAKNCP 918
Db 178 CVCPQGLP-----PDLKRGNTS-----TVEFKQPC---FLEISADHPG 216

QY 919 LSNPKNDGTNSDPVDFYRCITPYFGKQGDV---PIHACISWPKHGGTCHLAGBED 976
Db 217 RSMYTCVQFGDKAS--SKQCQPGYHSGTCELLLEKEDCSANPCSH--GVCISFG-- 269

QY 977 GPWICADGPEGENGCVNVYDCCDNDCENNSTCVGINNYTLCPPETGELCEKLDPC 1036
Db 270 GPQICDDGYSGSYCGGKGNVNNCKAGSKGNGVNSYFDCPCPERTGPI--KMD-C 327

QY 1037 AQDLNPGQDHSKCI---LTPKGFKCDTPGVYGHCDIDFDC-QDNKCNAGHCTDAVN 1092
Db 328 SAIPDCIMGT--CDSPLSKAFQCGRPYEGILCGQDNKLESEMCLNNGTCLNPLG 386

QY 1093 GTYICPEGYSLGFCPEPWWLPRTSPCDNFQDQAGCI-VRIPEQICQLPGYQEK 1151
Db 387 SPFRCDCARGFGKWD--EPL-----NCCDFHCENDTCHMTSDHSPVQCKNPGTGR 439

QY 1152 CEKLVSVNFINKESYLQPSAKVRPTNITQIANDSDGILLYKGDKHIAVELTRGV 1211
Db 440 CEKECPIGF-----GGVR--CDLARE---GICSRQGG----- 468

QY 1212 RASTDYGSPHPSAYSYSETINGNHYVELLALDQSLSLSDVGGNPKIITNLKSQSTLNF 1271
Db 469 -----CFNQG--KCLSGFC-----Y 481

```

Qy 1272 DSFLYGVGMPKSNVASLRQAPGNGTSFHCGRINLYINSELQDFQKVPMTGILGPCP 1331
 Db 482 CPDPFGT-----NQCENKNGKSSLSLENL-----SDP 511

Qy 1332 CHKKVCAHGT-CQSSQAGFTCECGWMPGLDQRTNDPCLNGKCHGTCLPNAFTS 1390
 Db 512 CMM-----NATCIDVDHATGYACICKQGFBDGIC-ERKDLCLNPNCSNGVCWQHREPS 566

Qy 1391 CKCLRGHGVCIDDEEDLFNCPQAIKRGKRLSLG-----GQPYCESSGTYGDCS 1443
 Db 567 CDDPGPGYGNCRQE-----KMFRCILSCQNGVCINDEEKRG-CESYGFSGARC 618

Qy 1444 DREISCRGERIDTYQKQGTACQTTKRVSL-----ECRGGC-----AGGQCCPLRS 1493
 Db 619 EKXINLGTETKDSLLR-----SVCEKRCESRANGCNDACNTAACKPGDGCSSG---- 670

Qy 1494 KRRKYSFECTDGGSFVD 1510
 Db 671 KREPPS-KRYTGNMAD 686

RESULT 14

FBP3_STRPU
 ID FBP3_STRPU STANDARD; PRT; 570 AA.
 AC P49013;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE FIBROPELLIN C PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 3)
 DE (EFG III) (FIBROPELLIN III).
 GN EGF3.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 RN [1]
 RF SEQUENCE FROM N.A.
 RC TISSUE=GASTROLA;
 RX MEDLINE=93273088; PubMed=8500658;
 RA Blagrove B.W., Raff R.A.;
 RT *The SpzP III gene encodes a member of the fibropellins: EGF repeat-
 containing proteins that form the apical lamina of the sea urchin
 embryo.*;
 RL Dev. Biol. 157:526-538(1993).
 CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
 MATRIX.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
 CC EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
 CC MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
 CC THROUGH SUBSEQUENT STAGES.
 CC -1- MISCELLANEOUS: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
 CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
 CC TO AVIDIN/STREPTAVIDIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L07045; AAA30045.1; ..
 DR HSP: P00740; IIXA.
 DR INTERPRO: IPR000088; ..
 DR INTERPRO: IPR000152; ..
 DR INTERPRO: IPR000091; ..
 DR INTERPRO: IPR000859; ..
 DR INTERPRO: IPR001438; ..

DR INTERPRO: IPR001881; ..
 DR PFAM: PF01382; Avidin; 1.
 DR PFAM: PF00431; CUB; 1.
 DR PFAM: PF00008; EGF; 8.
 DR PRINTS: P00010; EGFBLD00.
 DR PROSITE: PS00010; ASX_HYDROXYL; 8.
 DR PROSITE: PS00022; EGF_1; 8.
 DR PROSITE: PS00577; AVIDIN; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR PROSITE: PS01185; EGF_2; 7.
 DR PROSITE: PS01187; EGF_Ca; 6.
 KW Biotin; EGF-like domain; Repeat; Signal; Glycoprotein.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 570 FIBROPELLIN C.
 FT DOMAIN 18 55 EGF-LIKE 1.
 FT DOMAIN 62 175 CUB.
 FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 366 402 EGF-LIKE 7.
 FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 442 570 AVIDIN-LIKE.
 FT DISULFID 23 34 BY SIMILARITY.
 FT DISULFID 28 43 BY SIMILARITY.
 FT DISULFID 45 54 BY SIMILARITY.
 FT DISULFID 180 191 BY SIMILARITY.
 FT DISULFID 185 200 BY SIMILARITY.
 FT DISULFID 202 211 BY SIMILARITY.
 FT DISULFID 218 229 BY SIMILARITY.
 FT DISULFID 223 238 BY SIMILARITY.
 FT DISULFID 240 249 BY SIMILARITY.
 FT DISULFID 256 267 BY SIMILARITY.
 FT DISULFID 261 276 BY SIMILARITY.
 FT DISULFID 278 287 BY SIMILARITY.
 FT DISULFID 294 305 BY SIMILARITY.
 FT DISULFID 299 314 BY SIMILARITY.
 FT DISULFID 316 325 BY SIMILARITY.
 FT DISULFID 332 343 BY SIMILARITY.
 FT DISULFID 337 352 BY SIMILARITY.
 FT DISULFID 354 363 BY SIMILARITY.
 FT DISULFID 370 381 BY SIMILARITY.
 FT DISULFID 375 390 BY SIMILARITY.
 FT DISULFID 392 401 BY SIMILARITY.
 FT DISULFID 408 419 BY SIMILARITY.
 FT DISULFID 413 428 BY SIMILARITY.
 FT DISULFID 430 439 BY SIMILARITY.
 FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 570 AA; 61116 MW; B665E3E1C05E5E CRC64;

Query Match 6.6%; Score 549.5; DB 1; Length 570;
 Best Local Similarity 39.7%; Pred. No. 3.2e-27;
 Matches 96; Conservative 33; Mismatches 98; Indels 15; Gaps 5;

Qy 918 CLSNPKNGKNGTNSDPVDFYRCTCPYGFQKQCDVPHACISNPKHGKGTCLAKGREDG 977
 Db 180 CTNPNCLMGATC-VQWVDYQICARGPTGNCRTIDDCASAPCRNGGAC---VQWVQ 235

Qy 978 FWCACIDGFEZGECYVWVDCEDNDCEBNSTCVGVNNTYCLPPTGTGLCEERLDPCA 1037
 Db 236 YTCNICPGFVGNKNNINDEASIPCLNGGICVGVNIPACTLPGVTGLICETIDNCA 295

Qy 1038 QLNPNKSGLCFPCSPFWLPTSPDNFQNGAGCIVRINEPICQLPGYQKGECKLWS 1157
 Db 296 S-SPOQMGSGCTDANVHYTCCDRAGFTGSGNCTNINEASSPCLNGSGLDGQVQVQ 353

Qy 1098 CPFGYSGLFCFPCSPFWLPTSPDNFQNGAGCIVRINEPICQLPGYQKGECKLWS 1157
 Db 354 CLNWTYTHGWSI-----LDACASLPCQNGVCTNNGVGVDCPLPGTGNCE--ID 404

Qy 1458 VN 1159

Db 405 IN 406

RESULT 15

GLP1_CAEEL

ID GLP1_CAEEL STANDARD; PRT; 1295 AA.

AC P31508;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE GLP-1 PROTEIN PRECURSOR.

GN GLP-1 OR EMB-33 OR P02A9.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chordata; Rhabditida; Rhabditidae;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RY MEDLINE=89336787; PubMed=2758466;

RA Tochen J., Greenwald L.;

RT "glp-1 and lin-12, genes implicated in distinct cell-cell

RT interactions in C. elegans, encode similar transmembrane proteins.";

RL Cell 58:553-563(1989).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RY MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,

RA Sims M., Smales N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,

RA Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [3]

RP DELETION OF 1174-1295.

RY MEDLINE=91351288; PubMed=1881436;

RA Mango S.E., Maine B.M., Kimble J.;

RT "Carboxy-terminal truncation activates glp-1 protein to specify

RT vulval fates in Caenorhabditis elegans.";

RL Nature 352:811-815(1991).

RN [4]

RP CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.

RY MEDLINE=93354444; PubMed=8350921;

RA Roehl H., Kimble J.;

RT "Control of cell fate in C. elegans by a GLP-1 peptide consisting

RT primarily of ankyrin repeats.";

RL Nature 364:632-635(1993).

RN [5]

RP FUNCTION.

RY MEDLINE=94208066; PubMed=8156602;

RA Nello C.C., Draper B.W., Pries J.R.;

RT "The maternal genes apx-1 and glp-1 and establishment of

RT dorsal-ventral polarity in the early C. elegans embryo.";

RL Cell 77:95-106(1994).

CC -1- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE

CC BLASTOMERES, ABA AND APA. PROPER SIGNALING BY GLP-1 INDUCES ABA

CC DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA

CC DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE

CC ESTABLISHMENT OF THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- DEVELOPMENTAL STAGE: ACTS ON ABA DEVELOPMENT DURING 4-CELL AND

CC 12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL

CC STAGES.

CC -1- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.

CC -1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.

CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M25580; AAA28058.1; .

DR EMBL: J19555; CAA79620.1; .

DR EMBL: J29116; CAA79620.1; JOINED.

DR EMBL: J29116; CAA82373.1; .

DR EMBL: J19555; CAA82373.1; JOINED.

DR PIR: A32901; A32901.

DR HSP: P00740; 1.IXA.

DR WORMPEP: P02A9.6; C00237.

DR INTERPRO: IPR000152; .

DR INTERPRO: IPR000051; .

DR INTERPRO: IPR000080; .

DR INTERPRO: IPR000181; .

DR INTERPRO: IPR000211; .

DR PFM: PFM00008; EGF; 10.

DR PFM: PFM00023; ank; 4.

DR PFM: PFM00066; notch; 3.

DR PROSITE: PS50088; ANK_REPEAT; 3.

DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE: PS00010; ASX_HYDROLYTIC; 2.

DR PROSITE: PS00022; EGF_1; 10.

DR PROSITE: PS01186; EGF_2; 8.

DR PROSITE: PS01187; EGF_CA; 1.

DR Differentiation: Repeat; ANK repeat; EGF-like domain; Transmembrane;

KW Glycoprotein; Signal.

FT SIGNAL 1 15 POTENTIAL.

FT CHAIN 16 1295 GLP-1 PROTEIN.

FT DOMAIN 16 764 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 765 786 POTENTIAL.

FT DOMAIN 787 1295 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 493 607 3 X LIN/NOTCH REPEATS.

FT DOMAIN 988 1133 6 X ANK MOTIF REPEATS.

FT DOMAIN 19 58 EGF-LIKE 1.

FT DOMAIN 117 152 EGF-LIKE 2.

FT DOMAIN 154 190 EGF-LIKE 3.

FT DOMAIN 190 230 EGF-LIKE 4.

FT DOMAIN 232 269 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 271 308 EGF-LIKE 6.

FT DOMAIN 316 359 EGF-LIKE 7.

FT DOMAIN 369 406 EGF-LIKE 8.

FT DOMAIN 407 443 EGF-LIKE 9.

FT DOMAIN 446 479 EGF-LIKE 10.

FT REPEAT 493 527 LIN/NOTCH 1.

FT REPEAT 528 568 LIN/NOTCH 2.

FT REPEAT 569 608 LIN/NOTCH 3.

FT REPEAT 915 946 ANK MOTIF 1.

FT REPEAT 947 987 ANK MOTIF 2.

FT REPEAT 988 1019 ANK MOTIF 3.

FT REPEAT 1020 1056 ANK MOTIF 4.

FT REPEAT 1057 1098 ANK MOTIF 5.

FT REPEAT 1099 1133 ANK MOTIF 6.

FT DISULFID 23 35 BY SIMILARITY.

FT DISULFID 29 46 BY SIMILARITY.

FT DISULFID 48 57 BY SIMILARITY.

FT DISULFID 121 131 BY SIMILARITY.

FT DISULFID 126 140 BY SIMILARITY.

FT DISULFID 142 151 BY SIMILARITY.

FT DISULFID 158 169 BY SIMILARITY.

FT DISULFID 163 178 BY SIMILARITY.

FT DISULFID 180 189 BY SIMILARITY.

```

QY 1197 GDAHDELAVELTGRGVASITDQSGHASAIYSVETIND-----NHWIHELVALDQSLSLGV 1252
Db 326 -----RPHCRSRNST-----CYNDGRGRCNGPCVCE----- 350

QY 1253 DGGNPKIITLNLSKSTLNDSPLTVGGMPGKSVASLRQAQPGQTSFAGCIENLYNSE 1312
Db 351 -----PDYIGER-----CEIN----- 361

QY 1313 LQDFQKFMQGTGILPGCEPCHKVKVA-HGTGQPSQAGFTCEQDEGWMPLCDITNDCP 1371
Db 362 RDKDF-----FPDQSGCKYPCVNNATIDILKNSGYSCHCLPGLYGLNCKEOL--LC 411

QY 1372 LGMKCVH-GTCLFINAFTYSCKCLEGGVGLCDEBEDLNNPQAKIKCB-HGCRSLSLGQ 1429
Db 412 TPTTCANGGTGCBGVWG-IRCPNCPNGSPSDCYEIKRDL--CSHPHCKMGVCKYGTG-- 465

QY 1430 PYCRSSSYTGDSDCRSISCRGER---IRDYQKQOQYAAQ--TYKKWSLRLE----- 1479
Db 466 -YCBQGTGTGTPCEVLVIERKSKETVRDLCQKQ-----CNDIASNGICNPCNLECN 520

QY 1480 ---GGCAGGCCQPLRSKRRKTSPECT 1504
Db 520 -----

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:19:59 ; Search time 559.88 Seconds
(without alignments)
319.251 Million cell updates/sec

Title: US-09-540-245A-2
Perfect score: 8316
Sequence: 1 MRGVGWQMLSLSLGLAIL.....SSPFDVZKVKVKCTKCVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	8316	100.0	1525	4 Q9Y5Q7	Q9Y5Q7 homo sapien
2	8265	99.4	1529	4 Q94813	Q94813 homo sapien
3	8257	99.3	1521	4 Q95710	Q95710 homo sapien
4	8095	97.3	1521	11 Q9R1B9	Q9R1B9 mus musculus
5	5728	68.9	1523	11 Q88280	Q88280 rattus norv
6	5714.5	68.7	1523	4 Q75094	Q75094 homo sapien
7	5703	68.6	1523	11 Q9WVB4	Q9WVB4 mus musculus
8	5597	67.3	1534	4 Q75093	Q75093 homo sapien
9	5578.5	67.1	1531	11 Q88279	Q88279 rattus norv
10	5538.5	66.6	1531	11 Q9WVB5	Q9WVB5 mus musculus
11	5530	66.5	1025	11 Q92166	Q92166 mus musculus
12	5497	66.1	1530	11 Q9WVC5	Q9WVC5 rattus norv
13	3934	47.3	796	11 Q9WVC1	Q9WVC1 rattus norv
14	3588	43.1	1504	5 Q9V7F9	Q9V7F9 drosophila
15	3586	43.1	1504	5 Q9XYV4	Q9XYV4 drosophila
16	3517	42.3	1480	5 Q9V7F8	Q9V7F8 drosophila
17	3071	36.9	850	4 Q95804	Q95804 homo sapien
18	2703.5	32.5	1440	5 Q20204	Q20204 caenorhabditis
19	2282.5	27.4	664	4 Q9U117	Q9U117 homo sapien

20	1318	15.8	333	4 Q9U7H5	Q9U7H5 homo sapien
21	1115.5	13.4	530	5 Q24526	Q24526 drosophila
22	818.5	9.8	2704	5 Q97458	Q97458 drosophila
23	818	9.8	2634	5 Q9W4T8	Q9W4T8 drosophila
24	790	9.5	2653	5 Q25253	Q25253 lucilia cup
25	768	9.2	1218	11 Q9QXX0	Q9QXX0 mus musculus
26	768	9.2	1219	11 Q63722	Q63722 rattus norv
27	766	9.2	1218	4 Q15816	Q15816 homo sapien
28	766	9.2	1218	4 Q14902	Q14902 homo sapien
29	766	9.2	1227	4 P78504	P78504 homo sapien
30	761	9.2	1218	4 Q15122	Q15122 homo sapien
31	757	9.1	2352	5 Q61240	Q61240 halocynthia
32	748.5	9.0	2447	13 Q13149	Q13149 figu rubrip
33	736.5	8.9	1193	13 Q90819	Q90819 gallus gall
34	736	8.9	1203	11 Q06008	Q06008 mus musculus
35	735	8.8	2470	11 Q35516	Q35516 mus musculus
36	732	8.8	2471	11 Q9QW30	Q9QW30 rattus norv
37	731	8.8	2531	5 Q16004	Q16004 lytechinus
38	719.5	8.7	2281	4 Q9UPL3	Q9UPL3 homo sapien
39	719.5	8.7	2321	4 Q9Y618	Q9Y618 homo sapien
40	719.5	8.7	2321	4 Q9UM47	Q9UM47 homo sapien
41	713.5	8.6	1212	13 Q42347	Q42347 gallus gall
42	711.5	8.6	2319	11 Q9R172	Q9R172 rattus norv
43	697	8.4	1254	13 Q9YHU2	Q9YHU2 brachydanio
44	691.5	8.3	1964	11 Q35442	Q35442 mus musculus
45	682.5	8.2	1238	4 Q9Y219	Q9Y219 homo sapien

ALIGNMENTS

RESULT 1
Q9Y5Q7
ID Q9Y5Q7 PRELIMINARY; PRT: 1525 AA.
AC Q9Y5Q7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLT1.
GN SLT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99200389; PubMed=10102266;
RA Wang K.H., Brose K., Arnott D., Kidd T., Goodman C.S., Henzel W.,
RA Tessier-Lavigne M.;
RT 'Biochemical purification of a mammalian slit protein as a positive
RT regulator of sensory axon elongation and branching.';
RL Cell 96:771-784(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT 'Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.';
RL Cell 96:795-806(1999).
DR EMBL: AF133270; AAD2539.1; .
DR HSP: P00743; 1CCP.
DR INTERPRO: IPR000152; .
DR INTERPRO: IPR000359; .
DR INTERPRO: IPR000372; .
DR INTERPRO: IPR000483; .
DR INTERPRO: IPR000561; .
DR INTERPRO: IPR000742; .
DR INTERPRO: IPR001611; .
DR INTERPRO: IPR001791; .
DR INTERPRO: IPR001881; .

DR INTERPRO: IPR002272; -.
 DR PFAM: PF00008; BGF; 9.
 DR PFAM: PF00054; laminin_6; 1.
 DR PFAM: PF00560; LRR; 18.
 DR PFAM: PF01462; LRRM; 4.
 DR PFAM: PF01463; LRRCT; 4.
 DR PRINTS: PR01143; FSHRECEPTOR.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.
 DR PROSITE: PS00022; BGF_1; UNKNOWN_9.
 DR PROSITE: PS01185; CCK_1; UNKNOWN_1.
 DR PROSITE: PS01186; BGF_2; 7.
 DR PROSITE: PS01187; BGF_CA; 2.
 DR PROSITE: PS01225; CCK_2; 1.
 KW Glycoprotein; BGF-like domain.
 SQ SEQUENCE 1525 AA; 169394 MW; 8A81CDE34EF06A73 CRC64;

Query Match 100.04; Score 8316; DB 4; Length 1525;
 Best Local Similarity 100.04; Pred. No. 0;
 Matches 1525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRGVMQMSLSGLVLAAILNKVAPQACPAQSCSSSTVDCGSLALSVPMNPRTEL 60
 1 MRGVMQMSLSGLVLAAILNKVAPQACPAQSCSSSTVDCGSLALSVPMNPRTEL 60
 61 DLGNMNTIRITITDPAGLRLHLVQLWENKISTIERGAFQDKELERLRLNHLQLPPE 120
 61 DLGNMNTIRITITDPAGLRLHLVQLWENKISTIERGAFQDKELERLRLNHLQLPPE 120
 121 LFLGLAKLYRLDLSNQIQAIPRAFRGVDINKLQDYNQISCIEDGAFRLADLEV 180
 121 LFLGLAKLYRLDLSNQIQAIPRAFRGVDINKLQDYNQISCIEDGAFRLADLEV 180
 181 FLNNNNITPLSVASFMWPKLATPLRLSNVLYDCCHLALDGLKLRPRVGLTYCQMP 240
 181 FLNNNNITPLSVASFMWPKLATPLRLSNVLYDCCHLALDGLKLRPRVGLTYCQMP 240
 241 HLAGHVAEVCQKREPVCSDEEGHQSFMAPSCSYLACPACTCSNNIVDCRGKLEIPT 300
 241 HLAGHVAEVCQKREPVCSDEEGHQSFMAPSCSYLACPACTCSNNIVDCRGKLEIPT 300
 301 NLPEITELRLQNTIKVIPPGAFSPYKLLRLDLSNQISELAPDQGLRLSLVLY 360
 301 NLPEITELRLQNTIKVIPPGAFSPYKLLRLDLSNQISELAPDQGLRLSLVLY 360
 361 GNKITELPLSPFGLPSQLLLNANKINCLRVDAFDQGLNLLSLDYNKLTITANGFT 420
 361 GNKITELPLSPFGLPSQLLLNANKINCLRVDAFDQGLNLLSLDYNKLTITANGFT 420
 421 SPLRAITQYMLAQNPFTCDCHLKLADYLRNITETSGARCTSPRLNKLRIQIKSKF 480
 421 SPLRAITQYMLAQNPFTCDCHLKLADYLRNITETSGARCTSPRLNKLRIQIKSKF 480
 481 RCGSTEDYRSLSGDCPADLACEPKRCBCTVDCSQNKLNIPEHPTQTAELRLNNE 540
 481 RCGSTEDYRSLSGDCPADLACEPKRCBCTVDCSQNKLNIPEHPTQTAELRLNNE 540
 541 PTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGNEILLTNRLENVQRMKFG 600
 541 PTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGNEILLTNRLENVQRMKFG 600
 601 LESLKTMLRSNRTICVNGDSFGLSVRLSLDYNQITTVAGAFDTLSLSTLNLAN 660
 601 LESLKTMLRSNRTICVNGDSFGLSVRLSLDYNQITTVAGAFDTLSLSTLNLAN 660
 661 PFNVCNLAWEMLRKKRIVTGNPCQKPYFLKEPIQDVAIQDFTCDGNDNDCSPL 720
 661 PFNVCNLAWEMLRKKRIVTGNPCQKPYFLKEPIQDVAIQDFTCDGNDNDCSPL 720
 721 SRCPECTCLDTVVRCSNGLKLVLPKIPRDTVELYLDGQFTPLVPELSNKHLYLID 780
 721 SRCPECTCLDTVVRCSNGLKLVLPKIPRDTVELYLDGQFTPLVPELSNKHLYLID 780

QY 781 SNNRISTLSNQSFSNNQTLTLLSYNLRICIPPTFDGLSKLSLHSGNDISVWPGA 840
 Db 781 SNNRISTLSNQSFSNNQTLTLLSYNLRICIPPTFDGLSKLSLHSGNDISVWPGA 840
 QY 841 FNDLSLSHSAIGANPLYCDNMQLSDWVSKYKEPIARCAAGPMGAKILLTPSKK 900
 Db 841 FNDLSLSHSAIGANPLYCDNMQLSDWVSKYKEPIARCAAGPMGAKILLTPSKK 900
 QY 901 PTCQGPVWNILAKNCPCLSNPKNDGTCSNDPVDYPTCTCYCFGQDCDVPBACIS 960
 Db 901 PTCQGPVWNILAKNCPCLSNPKNDGTCSNDPVDYPTCTCYCFGQDCDVPBACIS 960
 QY 961 PCKHGCTCHLAGEEDGFWCICADGFEGENCEVNVDCENDCENNSTCVGILNWTCLC 1020
 Db 961 PCKHGCTCHLAGEEDGFWCICADGFEGENCEVNVDCENDCENNSTCVGILNWTCLC 1020
 QY 1021 PPEYTGELCEKILDFCAQLNCPQDHSKCLITPKGFKCDCTPGYVGEHCDIDFDCQDNK 1080
 Db 1021 PPEYTGELCEKILDFCAQLNCPQDHSKCLITPKGFKCDCTPGYVGEHCDIDFDCQDNK 1080
 QY 1081 CKMGACTDVAWGVTCTICPEGSLGCFEFSPPMVLPRSTPCDNDCQNGACIVIRNEPI 1140
 Db 1081 CKMGACTDVAWGVTCTICPEGSLGCFEFSPPMVLPRSTPCDNDCQNGACIVIRNEPI 1140
 QY 1141 CQCPLGCTGCHLGVSVWPKINKEYSYLIQSAKVRPQNTITLQATDEDSGLLYKGDM 1200
 Db 1141 CQCPLGCTGCHLGVSVWPKINKEYSYLIQSAKVRPQNTITLQATDEDSGLLYKGDM 1200
 QY 1201 HIAVELYAGRVASDYTGSHPASISYVETINDGNHVIYELLADQSLSLSVGQGNPKI 1260
 Db 1201 HIAVELYAGRVASDYTGSHPASISYVETINDGNHVIYELLADQSLSLSVGQGNPKI 1260
 QY 1261 TNLKQSTLWDSPLTVGCMGKSNWASLQAPQNGTSFSGCRNLNINSELQDQKVP 1320
 Db 1261 TNLKQSTLWDSPLTVGCMGKSNWASLQAPQNGTSFSGCRNLNINSELQDQKVP 1320
 QY 1321 MQGILPCECHKVKCAEGTQCSQSQGFTCEQCSQSGWMLCDQRTNDPCLGNKVCWT 1380
 Db 1321 MQGILPCECHKVKCAEGTQCSQSQGFTCEQCSQSGWMLCDQRTNDPCLGNKVCWT 1380
 QY 1381 CLPINFATSYCKCLEGHGVLCDEEDLFNQCIAIKHGKRLSLGLQPTCCSGTGT 1440
 Db 1381 CLPINFATSYCKCLEGHGVLCDEEDLFNQCIAIKHGKRLSLGLQPTCCSGTGT 1440
 QY 1441 DSCDREISCRGERIDTYQKQGYAACQITKKVSRLECGGCAGGCGGCPPLSKRKYSP 1500
 Db 1441 DSCDREISCRGERIDTYQKQGYAACQITKKVSRLECGGCAGGCGGCPPLSKRKYSP 1500
 QY 1501 ECTDGSSTFDEVERVKCGTCRCVS 1525
 Db 1501 ECTDGSSTFDEVERVKCGTCRCVS 1525

RESULT 2

ID 094813 PRELIMINARY; PRT; 1529 AA.
 AC 094813;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SLIT-2 PROTEIN.
 GN SLIT-2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99033071; PubMed=9813312;
 RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;
 RT "Cloning and expressions of three mammalian homologues of Drosophila
 RT slit suggest possible roles for Slit in the formation and maintenance
 RT of the nervous system.";

RU Brain Res. Mol. Brain Res. 62:175-186(1998).

DR EMBL: AB017168; BAA35185.1; -.

DR HSP: P00743; ICCF.

DR INTERPRO: IP000152; -.

DR INTERPRO: IP000359; -.

DR INTERPRO: IP000372; -.

DR INTERPRO: IP000483; -.

DR INTERPRO: IP000561; -.

DR INTERPRO: IP000742; -.

DR INTERPRO: IP000611; -.

DR INTERPRO: IP0001791; -.

DR INTERPRO: IP0001881; -.

DR INTERPRO: IP0002272; -.

DR PFAM: PF00008; EGF; 9.

DR PFAM: PF00054; lamia.G.1.

DR PFAM: PF00560; LRR; 18.

DR PFAM: PF01462; LRRNT; 4.

DR PFAM: PF01463; LRRCT; 4.

PRINTS: PR01143; FSRECEPTOR.

PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.

DR PROSITE: PS00022; EGF_1; UNKNOWN_9.

DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.

DR PROSITE: PS01186; EGF_2; 7.

DR PROSITE: PS01187; EGF_CA; 2.

DR PROSITE: PS01225; CTCK_2; 1.

KW glycoprotein; EGF-like domain.

SQ SEQUENCE 1529 AA; 16986 MW; 5D19CC5E7FD461BA CRC64;

Query Match 99.48; Score 8265; DB 4; Length 1529;
Best Local Similarity 99.24; Pred. No. 0;
Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 1 MGVGQMLSLSLGLVLAIAIKVAPQAPACSCSGSTVCHGLALSVPRNPRMTERL 60

Db 1 MGVGQMLSLSLGLVLAIAIKVAPQAPACSCSGSTVCHGLALSVPRNPRMTERL 60

Qy 61 DLNNGNITRTITKDFAGLAHLVQLMKNKSTIERGAPQGLCEKLEKLNWNNHQLFPE 120

Db 61 DLNNGNITRTITKDFAGLAHLVQLMKNKSTIERGAPQGLCEKLEKLNWNNHQLFPE 120

Qy 121 LFLFGTAKLYRLDLSNMQIQAIPKAPFAGVDIRKLNQLDINQISCTEDGAPALRLDVL 180

Db 121 LFLFGTAKLYRLDLSNMQIQAIPKAPFAGVDIRKLNQLDINQISCTEDGAPALRLDVL 180

Qy 181 TLNNGNITRLSVASFNMPLKRTFRLHNSNLYCDCHLAWLSDWLRKRPVNGLYTQCMGPS 240

Db 181 TLNNGNITRLSVASFNMPLKRTFRLHNSNLYCDCHLAWLSDWLRKRPVNGLYTQCMGPS 240

Qy 241 HLHGHVAVQRFVFCDSDEEGHQSPMAPSCSLVHPCAACTCSNNIVDCRGKGLTEIPT 300

Db 241 HLHGHVAVQRFVFCDS---GHQSPMAPSCSLVHPCAACTCSNNIVDCRGKGLTEIPT 296

Qy 301 NLPTETITRLDQNTIKVIPPAGFSPYKALKRIDLSNNQISLAPADQGLRSLNSLVLY 360

Db 297 NLPTETITRLDQNTIKVIPPAGFSPYKALKRIDLSNNQISLAPADQGLRSLNSLVLY 356

Qy 361 GNKITEPLKSLFEGFLSLQLLLNANKINCILVDAFDQNLNLLNLTYNKLTQIANGTF 420

Db 357 GNKITEPLKSLFEGFLSLQLLLNANKINCILVDAFDQNLNLLNLTYNKLTQIANGTF 416

Qy 421 SPLRIAQTHLAQNPFITCDCHLAWLADYLTNPITSGARCTSPRLANRIGQIKSKF 480

Db 417 SPLRIAQTHLAQNPFITCDCHLAWLADYLTNPITSGARCTSPRLANRIGQIKSKF 476

Qy 481 RCS-----GTEDYRSKLSGDCFDIACPEKRCBGITVDCSNQNLNKPPIHPITQTTA 532

Db 477 RCSAKETPIPTGTDYRSKLSGDCFDIACPEKRCBGITVDCSNQNLNKPPIHPITQTTA 536

Qy 533 ELRLNNEFTVLEATGIFKPLQPKRKINFSNNKITDIEGAFGAGSGVNEILLTNSRLN 592

Db 537 ELRLNNEFTVLEATGIFKPLQPKRKINFSNNKITDIEGAFGAGSGVNEILLTNSRLN 596

Qy 593 VQRMKFGLESILKTLMLRSNRITCVGNSDFGLSSVRLLSLYNDQITTVAGAFDTLHSL 652

Db 597 VQRMKFGLESILKTLMLRSNRITCVGNSDFGLSSVRLLSLYNDQITTVAGAFDTLHSL 656

Qy 653 STNLNLANFPNCNYLAWLGENLKRKRITVGNPCQKPYFLKEIPQDAIQPTCDGNG 712

Db 657 STNLNLANFPNCNYLAWLGENLKRKRITVGNPCQKPYFLKEIPQDAIQPTCDGNG 716

Qy 713 DNGSCSPLSRPTCTCLDITVRCNKGKLVLPKGIPTDVTELYDGNQPTLVPKLSNY 772

Db 717 DNGSCSPLSRPTCTCTCLDITVRCNKGKLVLPKGIPTDVTELYDGNQPTLVPKLSNY 776

Qy 773 KHLTLIDLNSNRISTLSNQGSFNMQLLLTILSYNLRALCIPPTFDGLSKRLSLHNGD 832

Db 777 KHLTLIDLNSNRISTLSNQGSFNMQLLLTILSYNLRALCIPPTFDGLSKRLSLHNGD 836

Qy 833 ISVPPGAFNDLSALSLAIGANPLYCCDNQMSDWVKSSEKPGIARACGPGEMADKL 892

Db 837 ISVPPGAFNDLSALSLAIGANPLYCCDNQMSDWVKSSEKPGIARACGPGEMADKL 896

Qy 893 LLTTPSKKFTCPQVDVNIARCNPLSNPKMGTCNSDPVDFRCTCTPGFGQDCDV 952

Db 897 LLTTPSKKFTCPQVDVNIARCNPLSNPKMGTCNSDPVDFRCTCTPGFGQDCDV 956

Qy 953 PIHACISNPKHSGTCHLKESEDEGFWICADGFEKCEKENVNDQDNCNNSTCTVG 1012

Db 957 PIHACISNPKHSGTCHLKESEDEGFWICADGFEKCEKENVNDQDNCNNSTCTVG 1016

Qy 1013 INNTVLCPPETTGELCEKLDLFCQDLNPNQDRSCILTPKGFKCDCTPGVGEHCDID 1072

Db 1017 INNTVLCPPETTGELCEKLDLFCQDLNPNQDRSCILTPKGFKCDCTPGVGEHCDID 1076

Qy 1073 FDDCCNKKNGACCTDAVNGTICTCPBSYGLACEFSPMPLPRTSPCDNFQCGMAQC 1132

Db 1077 FDDCCNKKNGACCTDAVNGTICTCPBSYGLACEFSPMPLPRTSPCDNFQCGMAQC 1136

Qy 1133 IVRNIPNQCQLPGYQGEKCKLVSNVNEKSYQIPSAVRPQNTILQIATEDSGI 1192

Db 1137 IVRNIPNQCQLPGYQGEKCKLVSNVNEKSYQIPSAVRPQNTILQIATEDSGI 1196

Qy 1193 LLKKGDKHIAVLYRGRVASTYDGSHPASAIYSVETINDGNFHVIELALQSLSLV 1252

Db 1197 LLKKGDKHIAVLYRGRVASTYDGSHPASAIYSVETINDGNFHVIELALQSLSLV 1256

Qy 1253 DGNPKNITNLKSGTLNFDPSFLYGMGPKNVSLAQAPQNGTSPHCRINLTINSE 1312

Db 1257 DGNPKNITNLKSGTLNFDPSFLYGMGPKNVSLAQAPQNGTSPHCRINLTINSE 1316

Qy 1313 LQDQKPVPMQILGPCPCCHKVCAHGTCPQSQAQPTCECQSGWGLPCDQRTNDPL 1372

Db 1317 LQDQKPVPMQILGPCPCCHKVCAHGTCPQSQAQPTCECQSGWGLPCDQRTNDPL 1376

Qy 1373 GNKCVHGTCLPINAPFSYCKLSEHGGVLCDEEDLFPQAIKCHKNGKRLSLGQPTC 1432

Db 1377 GNKCVHGTCLPINAPFSYCKLSEHGGVLCDEEDLFPQAIKCHKNGKRLSLGQPTC 1436

Qy 1433 ECGSSGYTGDSCREISCRGERIDYIQKQGYAACOTTKVRSLRECAGGACGCGCPLR 1492

Db 1437 ECGSSGYTGDSCREISCRGERIDYIQKQGYAACOTTKVRSLRECAGGACGCGCPLR 1496

Qy 1493 SKRRKYSPECIDGSSFDEVEKVKCGTCRCVS 1525

Db 1497 SKRRKYSPECIDGSSFDEVEKVKCGTCRCVS 1529

RESULT 3

095710

ID 095710 PRELIMINARY; PRT; 1521 AA.

AC 095710;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE NEUROGENIC EXTRACELLULAR SLIT PROTEIN SLIT2.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homioidea; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 KC TISSUE=KIDNEY, BRAIN;
 MEDLINE=99279238; PubMed-10349621;
 RA Holmes G.P., Megus K., Burridge L., Raman S., Alger E., Yamada T.,
 Little M.H.;
 RT "Distinct but overlapping expression patterns of two vertebrate slit
 RT homologs implies functional roles in CNS development and
 RT organogenesis.";
 RL Mech. Dev. 79:57-72(1998).
 DR EMBL: AF055585; AAD04309.1; .
 DR HSP: P00743; ICCP.
 DR INTERPRO: IPR000152; .
 DR INTERPRO: IPR000359; .
 DR INTERPRO: IPR000372; .
 DR INTERPRO: IPR000483; .
 DR INTERPRO: IPR000561; .
 DR INTERPRO: IPR000742; .
 DR INTERPRO: IPR001611; .
 DR INTERPRO: IPR001791; .
 DR INTERPRO: IPR001881; .
 DR INTERPRO: IPR002272; .
 DR PFAM: PF00008; EGF_9.
 DR PFAM: PF00004; laminin_G.1.
 DR PFAM: PF00560; LRR_18.
 DR PFAM: PF01462; LRRNT_4.
 DR PFAM: PF01463; LRRCT_4.
 DR PRINTS: PR01143; PSRECEPTOR.
 DR PROSITE: PS00001; ASX_HYDROXYL; UNKNOWN_2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 7.
 DR PROSITE: PS01187; EGF_CA_2.
 DR PROSITE: PS01225; CTCK_2; 1.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1521 AA; 168947 MW; C05ADF7D78C48C9 CR64;

Query Match 99.3%; Score 8257; DB 4; Length 1521;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 1515; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

Qy 1 MRGVGMQLSLGLVLAINKVAPQACQSCSGSTVDCGLALRSVPNIPNTERL 60
 Db 1 MRGVGMQLSLGLVLAINKVAPQACQSCSGSTVDCGLALRSVPNIPNTERL 60
 61 DLNNGNITRTITDPAFLRLRLVLMENKISTIERGAFDLKELERLARNNHQLPFE 120
 Db 61 DLNNGNITRTITDPAFLRLRLVLMENKISTIERGAFDLKELERLARNNHQLPFE 120
 Qy 121 LFLGLTAKLRLDLSNQIAIPKAFAGAVDINQLNDQISCTEDGAFRALRDEVL 180
 Db 121 LFLGLTAKLRLDLSNQIAIPKAFAGAVDINQLNDQISCTEDGAFRALRDEVL 180
 Qy 181 TLNNNITLRSVAFNMPKRLTRFLHSNNLYCDCHLWLSDLRRKRPVGLTQCMGS 240
 Db 181 TLNNNITLRSVAFNMPKRLTRFLHSNNLYCDCHLWLSDLRRKRPVGLTQCMGS 240
 Qy 241 HLHGNVAEYQREFVCSDEEGBQSPMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIP 300
 Db 241 HLHGNVAEYQREFVCS---GQSPMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIP 296
 Qy 301 NLPEPTIETRLDQNTKIVIPGAFSPYKRLRIDLSNNQISLAPDAFQGLRSLSNLVLY 360
 Db 297 NLPEPTIETRLDQNTKIVIPGAFSPYKRLRIDLSNNQISLAPDAFQGLRSLSNLVLY 356
 Qy 361 GNKITLPKSLPEGLFSLQLLLNANKINCLRDVAPQDLHNNLLSLDNLKQITAKGT 420
 Db 357 GNKITLPKSLPEGLFSLQLLLNANKINCLRDVAPQDLHNNLLSLDNLKQITAKGT 416

Qy 421 SPLRAIQTHLAQNPFI CDCHLWLAIDLHNTPIETSGARCTSPRLANKRIQIISKKF 480
 Db 417 SPLRAIQTHLAQNPFI CDCHLWLAIDLHNTPIETSGARCTSPRLANKRIQIISKKF 476
 Qy 481 RCGSTEDYKSLSGDCFDALACPEKRCBGSTVDCSNQNLNKEPIHPIQTAELRLNNE 540
 Db 477 RCGSTEDYKSLSGDCFDALACPEKRCBGSTVDCSNQNLNKEPIHPIQTAELRLNNE 536
 Qy 541 PTVLTAETGFKLQPLAKINFSNNKTDIEGAFEGASGVNLLTSNRLNQHVMFKG 600
 Db 537 PTVLTAETGFKLQPLAKINFSNNKTDIEGAFEGASGVNLLTSNRLNQHVMFKG 596
 Qy 601 LESLATLMLRSNRTVCGNDSFGLSSVRLSLDNYQITVAPGAFDTLASLSLMLNLAN 660
 Db 597 LEXPMQLMLRSNRTVCGNDSFGLSSVRLSLDNYQITVAPGAFDTLASLSLMLNLAN 656
 Qy 661 PFNMCNCLAMLEWLKRRKRVGNPCOKPYFLKEIPIDVAIQDPTCDGNDGNSCPL 720
 Db 657 PFNMCNCLAMLEWLKRRKRVGNPCOKPYFLKEIPIDVAIQDPTCDGNDGNSCPL 716
 Qy 721 SRCPTCTCLTVVRCNKGKLVKPLGIPRODVTELDGQNTLPKLSNTHKLLIDL 780
 Db 717 SRCPTCTCLTVVRCNKGKLVKPLGIPRODVTELDGQNTLPKLSNTHKLLIDL 776
 Qy 781 SNKRISLTSNQSFSNMQLLLLSYNLRKCPHPTFGKSLALLSLHNDISVWPGEA 840
 Db 777 SNKRISLTSNQSFSNMQLLLLSYNLRKCPHPTFGKSLALLSLHNDISVWPGEA 836
 Qy 841 FNDLSALSHLAGANPLCYDCNQWLSNWSVEKEPGIARCAAGSGEMAKLLITPSSK 900
 Db 837 FNDLSALSHLAGANPLCYDCNQWLSNWSVEKEPGIARCAAGSGEMAKLLITPSSK 896
 Qy 901 FCTGCGVDNIIARCNFCLSNPKNDGTCSNDPVRCTCPGFGKQDCDVPFRACIS 960
 Db 897 FCTGCGVDNIIARCNFCLSNPKNDGTCSNDPVRCTCPGFGKQDCDVPFRACIS 956
 Qy 961 PKCGKGTCHLKEGEDGFWICADGFEBCYVNDCEDNCENNSTCVGINNNYTLIC 1020
 Db 957 PKCGKGTCHLKEGEDGFWICADGFEBCYVNDCEDNCENNSTCVGINNNYTLIC 1016
 Qy 1021 PPEYTGELCEKLDQFQALNPQKSDCILTPKPGKCDCTPGYVGHKIDFDQDQNK 1080
 Db 1017 PPEYTGELCEKLDQFQALNPQKSDCILTPKPGKCDCTPGYVGHKIDFDQDQNK 1076
 Qy 1081 CKNAGCATDAWNGTTCIPBEGSLCFSPSPMLPRTSPCDNFQCGMAQICVIRNEPI 1140
 Db 1077 CKNAGCATDAWNGTTCIPBEGSLCFSPSPMLPRTSPCDNFQCGMAQICVIRNEPI 1136
 Qy 1141 CQCLPGYQGECKELSVNFINESYQLIPSAKVRPTNITLIQATDSDGILLYKGKD 1200
 Db 1137 CQCLPGYQGECKELSVNFINESYQLIPSAKVRPTNITLIQATDSDGILLYKGKD 1196
 Qy 1201 HIAVELYGRVASTDTGSHASAIYSVETINDGNFHVILLALQSLSDVSGNKII 1260
 Db 1197 HIAVELYGRVASTDTGSHASAIYSVETINDGNFHVILLALQSLSDVSGNKII 1256
 Qy 1261 TNLKSGTLNDSFLYVGGMGKSNVASLRAPQNGTSPHGCINRLINSELDQKQVP 1320
 Db 1257 TNLKSGTLNDSFLYVGGMGKSNVASLRAPQNGTSPHGCINRLINSELDQKQVP 1316
 Qy 1321 MQTGLPGCECHKVKCAHGTCPSSQAGFTCEQBGWGLCDQRTDNPCLGNKVHGT 1380
 Db 1317 MQTGLPGCECHKVKCAHGTCPSSQAGFTCEQBGWGLCDQRTDNPCLGNKVHGT 1376
 Qy 1381 CLPNAFSTSCCKLBSGGVLCDEEDLFPNQAIKCKHCKSLGLQPTCECSGTYG 1440
 Db 1377 CLPNAFSTSCCKLBSGGVLCDEEDLFPNQAIKCKHCKSLGLQPTCECSGTYG 1436
 Qy 1441 DSCDREISCRGERIDTYQKQGYAACQTTKVSRLCEAGGACGGCQPLRSKRRTYSP 1500
 Db 1437 DSCDREISCRGERIDTYQKQGYAACQTTKVSRLCEAGGACGGCQPLRSKRRTYSP 1496

Qy 1501 ECTDSSSFVDEYKVKCGCTRCVS 1525
 |||||
 Db 1497 ECTDSSSFVDEYKVKCGCTRCVS 1521

RESULT 4
 Q9R189
 ID Q9R189 PRELIMINARY; PRT; 1521 AA.
 AC Q9R189;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SLIT2.
 GN SLIT2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER/ICR;
 RA Yuan W., Zhou L., Chen J.-h., Wu J.Y., Rao Y., Ornitz D.M.;
 RT "The mouse SLIT family: Secreted ligands for Robo expressed in
 patterns that suggest a role in morphogenesis and axon guidance.";
 RL Dev. Biol. 0:0-0(1999).
 DR EXBL: AFI44628; AAD4759.1; .
 DR HSP: P00743; I0CF.
 DR MGD: MGI:1315205; Slit2.
 DR INTERPRO: IP8000152; .
 DR INTERPRO: IP8000359; .
 DR INTERPRO: IP8000372; .
 DR INTERPRO: IP8000483; .
 DR INTERPRO: IP8000561; .
 DR INTERPRO: IP8000742; .
 DR INTERPRO: IP8001611; .
 DR INTERPRO: IP8001791; .
 DR INTERPRO: IP8001881; .
 DR INTERPRO: IP8002272; .
 DR PFM: PF00008; EGF_9.
 DR PFM: PF00054; Laminin_G; 1.
 DR PFM: PF00560; LRR; 18.
 DR PFM: PF01462; LRRNT; 4.
 DR PFM: PF01463; LRRCT; 4.
 DR PRINTS: PR00013; LEURICHRPT.
 DR PRINTS: PR01143; FSHRECEPTOR.
 DR PROSITE: PS00010; ASX_HIDROXIL; UNKNOWN_2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE: PS01185; CTRK_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 7.
 DR PROSITE: PS01187; EGF_CA; 2.
 SQ SEQUENCE 1521 AA; 168769 MW; 97DCA361578978EA CRC64;

Query Match 97.3%; Score 8095; DB 11; Length 1521;
 Best local similarity 95.5%; Pred. No. 0;
 Matches 1471; Conservative 34; Mismatches 16; Idels 4; Gaps 1;

Qy 1 MRCVGMQLSLGLVLAILNKVAPQAPCAQCSGSTVDCGSLALSVPNIPRTERL 60
 |||||
 Db 1 MRCVGMQLSLGLVLAILNKVAPQAPCAQCSGSTVDCGSLALSVPNIPRTERL 60
 |||||
 Qy 61 DLNGNITRITKTDPAGLRHLRLVLMENKSTIERGAFQDLKELRLNRRNLQLPPE 120
 |||||
 Db 61 DLNGNITRITKTDPAGLRHLRLVLMENKSTIERGAFQDLKELRLNRRNLQLPPE 120
 |||||
 Qy 121 LLEFGTAKLYALDLSNQIQAIPKAFRGADVINKLQDYQNSICIEGAFRALRDLEVL 180
 |||||
 Db 121 LLEFGTAKLYALDLSNQIQAIPKAFRGADVINKLQDYQNSICIEGAFRALRDLEVL 180
 |||||
 Qy 181 TLNNNITRLSVASFNMPKLRITFLHSNNLYCDCLAWLSDWLKRPVGLITQCMPS 240
 |||||
 Db 181 TLNNNITRLSVASFNMPKLRITFLHSNNLYCDCLAWLSDWLKRPVGLITQCMPS 240
 |||||

Qy 241 HLGRHNAEVRQREFVCSDEEGHQSPAPSCSVLHCPAECTSNNVIDCRGGLTETPT 300
 |||||
 Db 241 HLGRHNAEVRQREFVCS---GHGSPAPSCSVLHCPAECTSNNVIDCRGGLTETPT 296
 |||||
 Qy 301 NLPEITTEIRLEQNTIKVIPPQAFSPYKLRILDSNNQISELAPDAFQGLSLNSLVLY 360
 |||||
 Db 297 NLPEITTEIRLEQNSIRVIPPQAFSPYKLRILDSNNQISELAPDAFQGLSLNSLVLY 356
 |||||
 Qy 361 GNKITEPLSLFEGFSLQLLLNANKINCLRVADPQDLNLLSLYDNKLTQIAKGT 420
 |||||
 Db 357 GNKITEPLSLFEGFSLQLLLNANKINCLRVADPQDLNLLSLYDNKLTQIAKGT 416
 |||||
 Qy 421 SPLRLAIQTHLQONPFIICDCKLWADLYLHTNPIETSGARCTSPRLANKRIQIKSKFP 480
 |||||
 Db 417 SALRLAIQTHLQONPFIICDCKLWADLYLHTNPIETSGARCTSPRLANKRIQIKSKFP 476
 |||||
 Qy 481 RCGSTEDYRSLGSCFDADLACPKRCCEGTVDSCSNQKLNKIPHIPOITAEALNANNE 540
 |||||
 Db 477 RCGSTEDYRSLGSCFDADLACPKRCCEGTVDSCSNQKLNKIPHIPOITAEALNANNE 536
 |||||
 Qy 541 FTVLEATGIFPKPLQARKINFSNNKIIDIEGAFGASGVNELLTSNLEWQHKMFG 600
 |||||
 Db 537 FTVLEATGIFPKPLQARKINFSNNKIIDIEGAFGASGVNELLTSNLEWQHKMFG 596
 |||||
 Qy 601 LESLAIQLRSLNITCVGNSDFGLSVLLSYDQITTVAPAGFDSLRLSTLNLAN 660
 |||||
 Db 597 LESLAIQLRSLNITCVGNSDFGLSVLLSYDQITTVAPAGFDSLRLSTLNLAN 656
 |||||
 Qy 661 PFWNCYHLANGELWLRKRIVGNPCQKPYFLKEIPIQVAIQDFCTDGDGDDNSCSPL 720
 |||||
 Db 657 PFWNCYHLANGELWLRKRIVGNPCQKPYFLKEIPIQVAIQDFCTDGDGDDNSCSPL 716
 |||||
 Qy 721 SRCPTECTCLDTVWCNSKGLVLPKGIPIEDVTELYLDGQNTLVPKELSNYKHLIDL 780
 |||||
 Db 717 SRCPTECTCLDTVWCNSKGLVLPKGIPIEDVTELYLDGQNTLVPKELSNYKHLIDL 776
 |||||
 Qy 781 SNNRISLNSGFSNMQLTLTLISYNRLACIPPTFDGLSLALSLAGNDISVVPGE 840
 |||||
 Db 777 SNNRISLNSGFSNMQLTLTLISYNRLACIPPTFDGLSLALSLAGNDISVVPGE 836
 |||||
 Qy 841 FNDLSALSALIAGNLYLDCNQWLSQWSEYKPGIARACAGPGMAKLLITLPSKK 900
 |||||
 Db 837 FNDLSALSALIAGNLYLDCNQWLSQWSEYKPGIARACAGPGMAKLLITLPSKK 896
 |||||
 Qy 901 FTCQGVDMVILAACNPLSNPKNDGTCNDSDVPYRATCPYGFAGQCDVPICACIS 960
 |||||
 Db 897 FTCQGVDMVILAACNPLSNPKNDGTCNDSDVPYRATCPYGFAGQCDVPICACIS 956
 |||||
 Qy 961 PCKHGCTCHLREGEDEGFCICADGFEGENCEVNWDCEDNENNSTVDGGINNYTCL 1020
 |||||
 Db 957 PCKHGCTCHLREGEDEGFCICADGFEGENCEVNWDCEDNENNSTVDGGINNYTCL 1016
 |||||
 Qy 1021 PPETYGELCEKLFQCAQDLNFCQDSKCLITPKGFCDCPTGYVGEHCDDIFDQDNK 1080
 |||||
 Db 1017 PPETYGELCEKLFQCAQDLNFCQDSKCLITPKGFCDCPTGYVGEHCDDIFDQDNK 1076
 |||||
 Qy 1081 CKNGAKCTDAVNGTICPCPEYSGLCFSPVPLPRTSPCNDPCQNGAQCIIRINEPI 1140
 |||||
 Db 1077 CKNGAKCTDAVNGTICPCPEYSGLCFSPVPLPRTSPCNDPCQNGAQCIIRINEPI 1136
 |||||
 Qy 1141 CQCLPGTGKCEKLVSVNFINKESYLQIPSAKVPQNTITQIATDEDSGILLYGKD 1200
 |||||
 Db 1137 CQCLPGTGKCEKLVSVNFINKESYLQIPSAKVPQNTITQIATDEDSGILLYGKD 1196
 |||||
 Qy 1201 HIAVELYGRVASTDTGSHPSASIVSVETINDGNHIVELLADQSLSLVDGNGPRII 1260
 |||||
 Db 1197 HIAVELYGRVASTDTGSHPSASIVSVETINDGNHIVELLADQSLSLVDGNGPKVI 1256
 |||||
 Qy 1261 TNLKSGTLNDSPLVYGMGPKGNVSLAQPGQNTSFHGCIRNLYINSELQDFKVP 1320
 |||||
 Db 1257 TNLKSGTLNDSPLVYGMGPKGNVSLAQPGQNTSFHGCIRNLYINSELQDFKVP 1316
 |||||
 Qy 1321 MQTGLPCECHKVCABGTCCQSSGAGTCBCQBGWMLPQDRNTPCLNGKPHWT 1380
 |||||

Db 1317 MQTGILPGCECHKVCAGHMCQPSQSGFTCECEBHWGFLCDQRTNDPLCNKNCVST 1376
 Qy 1381 CLPINFATSCNCLBHGGLVDEEDLFNPOCMICKHKGKRLSGVQPCSCSGT 1440
 Db 1377 CLPINFATSCNCLBHGGLVDEEDLFNPOCMICKHKGKRLSGVQPCSCSGT 1436
 Qy 1441 DSCDREISCRGERINDYQKQGYACQTKKVSLECRGAGGCGGLSLKRRKST 1500
 Db 1437 DSCDREISCRGERINDYQKQGYACQTKKVSLECRGAGGCGGLSLKRRKST 1496
 Qy 1501 ECTDGSFVDEYEVKVKCGCTCVS 1525
 Db 1497 ECTDGSFVDEYEVKVKCGCTCVS 1521

RESULT 5

088280

ID 088280 PRELIMINARY; PRT: 1523 AA.

AC 088280;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE MGFS.

GN MGFS.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID:10116;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;

RY MEDLINE-98360089; PubMed-9693030;

RA Nakayama M., Wakajima D., Nagase T., Nomura N., Seki N., Ohara O.;

RT "Identification of high-molecular-weight proteins with multiple EGF-

like motifs by motif-trap screening.;"

RL Genomics 51:27-34(1998).

DR ENKL; AB011531; BAA32461.1; .

DR HSP; P01132; IEGF.

DR INTERPRO; IPR000152; .

DR INTERPRO; IPR000359; .

DR INTERPRO; IPR000372; .

DR INTERPRO; IPR000483; .

DR INTERPRO; IPR000561; .

DR INTERPRO; IPR000742; .

DR INTERPRO; IPR001611; .

DR INTERPRO; IPR001791; .

DR INTERPRO; IPR001881; .

DR INTERPRO; IPR002400; .

PFAM; PF00007; Cys_knot; 1.

PFAM; PF00008; EGF; 9.

PFAM; PF00054; laminin_G; 1.

PFAM; PF00560; LRR; 19.

PFAM; PF01462; LRRNT; 4.

PFAM; PF01463; LRRCT; 4.

PRINTS; PR00438; GFCYSKNOT.

DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.

DR PROSITE; PS00022; EGF_1; UNKNOWN_9.

DR PROSITE; PS01185; CTRK_1; UNKNOWN_1.

DR PROSITE; PS01186; EGF_2; 7.

DR PROSITE; PS01187; EGF_CA; 2.

DR PROSITE; PS01225; CTRK_2; 1.

KW Glycinein; EGF-like domain.

SQ SEQUENCE 1523 AA; 167766 MW; 6CE1B7AF244478B CRC64;

Query Match 68.9%; Score 5728; DB 11; Length 1523;
 Best Local Similarity 66.9%; Pred. No. 0;
 Matches 1015; Conservative 223; Mismatches 265; Indels 14; Gaps 9;

Qy 11 LSLGLVLA-ILNVAPQCAQPCSCSGSTVDCGLALRSVPNPRWTERDLGNNNIT 69
 Db 16 LALALALASILSGPPAAACPTKCTSAASVDCGLGLRVRGPRINRERDLDRNNIT 75

Qy 70 IYKTFAGLRLHLVQLMKNKISTIERGAFQDLKELERLRLNNHQLPELLFLGTLK 129

Db 76 IYKMDFTGLALRLVHLEDNQVSVIERGAFQDLKELERLRLNNHQLPELLFQSTPKL 135

Qy 130 YRLDLENGIQIAPKFAFGAVDIKNLDQNYISCIEDGAFRALRDLVLTNNNNIT 189

Db 136 YRLDLENGIQIAPKFAFGVGVKNLDNNNNITSCIEDGAFRALRDLVLTNNNNIT 195

Qy 190 LSVASFNPMKIRLTFRLSHNNLYCCHLAWLSDWLKRPVGLITQCGPSLRHGNVAE 249

Db 196 ILVTSFNPMKIRLRLSHNNLYCCHLAWLSDWLKQRRTIGQPTLCMAVPLRGFSVAD 255

Qy 250 VQKEHFVCSDEERHGSFMAPSCSV--LHCPACTCSNNIVDCRGKGLTEIPTNLPEIT 307

Db 256 VQKEHYVC---PGPHS-EAPACNANSLSCPSACSCNNIVDCRGKGLTEIPALPGIV 310

Qy 308 EIRLEQNTVITPPGAFSPYKRLRRIDLSNQISLEAPAFQGLSLNLSLVYGNKITEL 367

Db 311 EIRLEQNSIKIPAGPIQYKLRIDLSKQISDAPAFQGLSLTSLVLYGNKITET 370

Qy 368 PKSLPEGLSLQLLLNANKINCLVDAFQDLNANLNSLDNLQITAGFTSPRLATQ 427

Db 371 PKGLPEGLSLQLLLNANKINCLVDAFQDLNANLNSLDNLQITAGFTSPRLATQ 430

Qy 428 TWHLGNPFICDHLWADYIHLNPIETSGARTSPRLANKRIGIKSLFKRCSTED 487

Db 431 TLHLGNPFVDCDHLWADYIHLNPIETSGARTSPRLANKRISQIKSLFKRCSTED 490

Qy 488 YRSLSDGCFADLACPEKRCSTGVDCSSQKLNKIPEHPIQPTAEELNHNPFVLEAT 547

Db 491 YRHSSECFDVLCEKRCSTGVDCSSQKLSRPSHLPETDRLANDNDIATLEAT 550

Qy 548 GIKPKLQALRNINNNKIIDIEGAFGAGSVNELLTSNRLNENVGHMKVGLSLATL 607

Db 551 GIKPKLQALRNINNNKIIDIEGAFGAGSVNELLTSNRLNENVGHMKVGLSLATL 610

Qy 608 MLRNSNITCVNDSFGLSVRLSLSDYNTIYVAGAFDTLSLSLTNLLANPFWCNCT 667

Db 611 MLRNSNITCVNDSFGLSVRLSLSDYNTIYVAGAFDTLSLSLTNLLANPFWCNCH 670

Qy 668 LAMLEGLKKRIVTGNFRCQPFPLKEIPIQDAIQTCDGNDNNSCSPLRCPTEC 727

Db 671 LAMLEGLKKRIVTGNFRCQPFPLKEIPIQDAIQTCDGNDNNSCSPLRCPTEC 729

Qy 728 TCLDTVTRCSHGKLVKPGIPDVTLYLDGNGFTLPKLSNYKHLTLDLSNNRIST 787

Db 730 TCVETVTRCSHGKLTLPKGMFDVTLYLDGNGFTLPKLSNYKHLTLDLSNNISIM 789

Qy 788 LSNQSFNMTQLTLLSYNRLACIPPTFDGLASRLSLHGHNDISVVPFAGNDSAL 847

Db 790 LTNHTFSNMTQLTLLSYNRLACIPVHFGNLSRLVLTGHNDISVVPFAGNDSAL 849

Qy 848 SHLAIANPLDCCNQNQLSDWVSEYKEGIARCAGGEMAKILLTTPSKKTCQGPV 907

Db 850 SHLAIANPLDCCNQLSEWIKAGYKCEIARCAGGEMAKILLTTPSKKTCQGPV 909

Qy 908 DVNIALKNCPLSNPCNDGTGNSDPVYRCTCYFGKQCDVPHACISNPKHGT 967

Db 910 DIIVACNACLSLSPCKNNGTGDQVQVRCYCTCYFGKQCDVPHACISNPKHGT 969

Qy 968 CHLKEEGDEGFWICADGFECEGNEVNDCCEDNCCNNSTCVGINITCLCPYETGE 1027

Db 970 CHLSEHSGDFSCSCLFGFEGQCEINPDCCEDNCCNSATCVGINITACVPYETGE 1029

Qy 1028 LCEELDFCAQIANPCQHSKCLTPKFGKCDCTPGYVGEBCIDFDQCNKNGNAC 1087

Db 1030 LCEVDINPEQNALQCHAKISLDKFGCEVPSYGLCEITDNDCAVAKRCHGAQC 1089

Qy 1088 TDVANGTYICPEYSGLFCEFSPPWLPTSPCDNFDQCAQCTVIRNIPQCQLPV 1147

Db 1090 VDAVNGTYICPGFSGLFCEHPPVLLQTPCDNFDQCAQCTVIRNIPQCQLPV 1149

Query Match 68.7%; Score 5714.5; DB 4; Length 1523;
Best Local Similarity 66.6%; Pred. No. 0;
Matches 1017; Conservative 216; Mismatches 278; Indels 15; Gaps 2;

QY	3	GVGWMLSLISGLVILA-LINKYAPQACPAQCSGSGVDCHGLALRYSPNTPNTRLE	60
Db	7	GGAAYRRLALALASVLSGPPVAPCTCTCSAASVDCBGLRAVPGIPNRAEL	66
QY	61	DLGNKNITRITKDFAGLHRLVLQNKENKSTIERGAFQDLAEELRLARNHNLQFPE	120
Db	67	DLDRKNITRITKDFAGLANKLRLHLEDNQVSVIERGAFQDLAQLERLARNKNLQVLE	126
QY	121	LLPLFATRLVRLDLSNQIQPKAPKRGANDIKMLQDLVQWISCTEDGAPALRDLEVL	180
Db	127	LLPQSPFLTRLDLSNQIQGQPKAPRGITDVKMLDNNHISCTEDGAPALRDLEIL	186
QY	181	TLNNNTIRLSVASFNMWPKLRTLRASNMLVCCDCHLAKLSDWLRKRPVGLVTCMGFS	240
Db	187	TLNNNISRIULTVTFNMWPKLRLRLASNMLVCCDCHLAKLSDWLRQRTRVQQTFLCAPP	246
QY	241	HLRGHNVAEVPKREFTCSDEEBGSGPASFVCSV---LPAACCTCSNNIVDCRGKLTEI	298
Db	247	HLRGFNWADVKEEYCPAPHS-----EPSCSNMISCPSCSTNNIVDCRGKLMEI	301
QY	299	PTNLPETTEIIEHLQNTIKVIPPAPSPFKKLRLDLSNQISHLAPDAGLGSLSNLSV	358
Db	302	PANLPGEVIEIHLQNKSIATAPAGVTFKLRLDISKNSIDAPADAGLGSLSVLSV	361
QY	359	LIGNKTEIPLKSLFGLSFLQQLLNNANKINLRVDPQDLNMLNLSLDNKLQITAKG	418
Db	362	LIGNKTEIATGLPDGLVSLQQLLNNANKINLRVTPQDLNMLNLSLDNKLQITSGK	421
QY	419	TFSPRLAQITMELAQNPFCDCHLKNLADLYLNPETSPGARCTSPRRLNKRIGQISK	478
Db	422	LEAPLQISQTLCHLQNPFCVDCHLKNLADLQNPETISGARCSSPRLNKRISQISK	481
QY	479	KFKCSGSDYRKSGLSCDPAFLCAPEKRCBGSTVDCSNQKLRIPEHPOTYAEALRN	538
Db	482	KFKCSGSDYRSFSSCFMDLVCEPKRCBGSTVDCSNQKLRIPEHPOTVYDLARN	541
QY	539	NEFTVLEATGFPKLPLQRLINVSNNKTIIEGAFGAGSVNEILNLSLENLQHKMF	598
Db	542	NEVSYLEATGFPKLPLNLRILNNSNKIKVREGAFGAGSVQELMLTGNLEVBHGVF	601
QY	599	RGLESKLTLMRNSRNTITCVGNSDIFGSLSVRLSLSDYQITVTPAGFDTLSLSTLNL	658
Db	602	RGLSKLTLMRNSNLGCVNSDITFAGLSVRLSLSDYLRITITVTPAGFDTVLSLSTNL	661
QY	659	ANPFCNMCNLYLMLGEWLKRRIVTGNPRCQPFYFLKEIPIDVIAQDCTGDGNDNSCS	718
Db	662	SNPFCNMCNCHLAKGLWLKRRIVSNGPRCQPFPELKEIPIDVIAQDCTD-GDNESSCQ	720
QY	719	PLSRCPTCTCLDTVPCRSNKLGLVPGKIPIDVLELVDGNQFTVPLKSLNKHVLTIL	778
Db	721	LSRCPDCTCMETVTVPCRSNKLGLAPGKMPDVEITVLEGNHVTAPPELSAHLVLTIL	780

RESULT 6
075094
ID 075094 PRELIMINARY; PRT; 1523 AA.
AC 075094;
DT 01-NOV-1998 (TRENDEL. 08, Created)
DT 01-AUG-1999 (TRENDEL. 11, Last sequence update)
DT 01-OCT-2000 (TRENDEL. 15, Last annotation update)
DE SLIT-3 PROTEIN (MEGF5).
GN SLIT-3 OR MEGF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99033071; PubMed=9813312;
RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;
"Cloning and expressions of three mammalian homologues of Drosophila
slit suggest possible roles for Slit in the formation and maintenance
of the nervous system.";
RT Brain Res. Mol. Brain Res. 62:175-186(1998).
RN [2]
RP SEQUENCE OF 785-1523 FROM N.A.
RX TISSUE=BRAIN;
RC MEDLINE=98360089; PubMed=9690330;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
like motifs by motif-triplet screening.";
RL Genomics 51:27-34(1998).
DR EMBL: AB017169; dbA35138.1; -.
DR EMBL: AB011538; dbA32466.1; -.
DR HSP; P01132; 1EGF.
DR INTERPRO: IPR000152; .
DR INTERPRO: IPR000359; .
DR INTERPRO: IPR00372; .
DR INTERPRO: IPR00483; .
DR INTERPRO: IPR00561; .
DR INTERPRO: IPR00742; .
DR INTERPRO: IPR01611; .
DR INTERPRO: IPR01791; .
DR INTERPRO: IPR01881; .
DR INTERPRO: IPR02049; .
DR INTERPRO: IPR03006; .
DR Pfam: PF00007; Cvs_knot. 1.

```

RESULT 7
C99WB4
ID C99WB4 PRELIMINARY; PRT; 1523 AA.
AC C99WB4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT3 (FRAGMENT).
GN SLIT3.
OS Mus musculus (Mouse).
OC Bakyrota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STAIN-SWISS WEBSTER/ICR;
RA Yuan Y., Zhou L., Chen J.-H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse SLIT family: Secreted ligands for Robo expressed in
RT patterns that suggest a role in morphogenesis and axon guidance."

```


Db 431 TLRLAAGPFCDCCHLWLDYLDQNPITSGARCSPPRLAKRISQISKXKFCSGSD 490

Qy 488 TRSKLSGDCPADLACEKRCBGTTVDCSNQKLNKIPHIPTQTLRLANNNEPTVLEAT 547

Db 491 TRNFSSEPCMDLVCPEKRCBGTTVDCSNQKLRIPSHLPETDRLDNLNDISVLEAT 550

Qy 548 GIPFKLPQLRKNFSNNKITTIDIEGAFGASGVNELLTNSRLNQHVKHFGLESKLT 607

Db 551 GIPFKLPNLRKLNNSNRKEVREGAFGASVEMLLNGQLMEHGRFRLSGKLT 610

Qy 608 MLRSNRTICVGNDSFGLSSVRLSLYDQITTVAGAFDTLHSLSTINLLNFWNCMY 667

Db 611 MLRSNLCVSCVNDTFAGLSSVRLSLYDNRTITTPGATTLVSLSTINLLNFWNCMH 670

Qy 668 LAMLGWLEKKRIVTGNPCQKPFYKEIPQIDVAIQDPTCDGNDNDSCPLSRCTPC 727

Db 671 MAMLGWLEKKRIVTGNPCQKPFYKEIPQIDVAIQDPTCDGNESSQCLSPRCPEQ 729

Qy 728 TCIQITVYKSNGLKLPKIPROVTEYLDGQNPITVPEKLSNKHILIDLSNWRIST 787

Db 730 TCIVETVYKSNGLKLPKIPROVTEYLDGQNPITVPEKLSNKHILIDLSNWRIST 789

Qy 788 LMSQSFNSMQLLILSYNRLACIPPTFDGLSKRLSLGNGDISVPEGAFNOLSA 847

Db 790 LNTWTFNSMSHMTLILSYNRLACIPVHAFNGLSRVLTLDGNDISSVPEGSFNLDS 849

Qy 848 SHLALGANPLCDCNMWLSVDWVYKKEFGIARCAAGGEMADKLLTTPSKITPCQGPV 907

Db 850 SHLALGANPLCDCNMWLSVDWVYKKEFGIARCAAGGEMADKLLTTPSHPCQGPV 909

Qy 908 DWTILAKNPLCSNPKNDGTCSNPDVFTICTPYKQGCQDVPYHACISNPKHGGT 967

Db 910 DWTILAKNPLCSNPKNDGTCSNPDVFTICTPYKQGCQDVPYHACISNPKHGGT 969

Qy 968 CHLKEEDGFCWICADGFGSEENVDNDCEDNENNSTVDGNNITCLPPTTGE 1027

Db 970 CHLSEMDGFCSCPLGFGSEENVDNDCEDNENNSTVDGNNITCLPPTTGE 1029

Qy 1028 LCEKLDLFCQALNQCQKSLCTPKGFCQDTPYVGHCDIDFDQCNKCKGABC 1087

Db 1030 LCEVDLITVPMNMLQCHKACISLDGFCRCVPGYSKLCITNDQCAHCKRGAQC 1089

Qy 1088 TDAVNGTTCICPEYSGFLCEFSPPMWLPTSPDNQDQNGAQCIVRIPEIQCLQPGY 1147

Db 1090 VDEYNGTTCICPEYSGFLCEFSPPMWLPTSPDNQDQNGAQCIVRIPEIQCLQPGY 1149

Qy 1148 QGKCEKLVSYNFKNSTYLQIPSAKRVPOITNLQIATDESGILYKGDHIAVELY 1207

Db 1150 AGFRCLEKLVSYNFKNSTYLQIPSAKRVPOITNLQIATDESGILYKGDHIAVELY 1209

Qy 1208 RGRVRYSTDTGSHPASALYSVETINDGNFIVHLLAQDLSLSDVGNPKIITNLSQK 1267

Db 1210 QGHRVLYTSLSSPPTTVISVETVNDGQPHSVKLVNMQTLNVLVYKDGAFSLKGLQKP 1269

Qy 1268 TLNFDSPPLTVGPMGKSNVSLAQAQNGTSPHGCNINLYNSLQDFQVPMQT-GIL 1326

Db 1270 AVGNSPLTGLGIPSTGLSALRQADPLGSGFHEVRLNINLQDFKALPQSLGVS 1329

Qy 1327 PGCEPCHKVCAHSTQCPSSQAGTCECQEGWMLDQRTNDCLNKCCHVCTCLPINA 1386

Db 1330 PCKSC--TVCHGLGCLSVKEDSVVCECHGWTGPIQDQEARDCGLSCHRGTCH-ATG 1386

Qy 1387 FYSCKCLGHBGGVLDDEEDLFPQCAIKCKHGKRLSGLGQPCBSSGTTGDSQDRE 1446

Db 1387 DSTVCKACBGTGALCDQKNDASASAFKCHBQCHIHDRGEPQLTQCPGSHHREDE 1446

Qy 1447 ISCGERIRDVTYRQGTAACTTKRVSRLKRCGGAGGQCCPLASRRKRTSFTCDGS 1506

Db 1447 NPMGCBVIRHARLQDYASCATSKVPMZCRGGC-GSQCCQPIKSRKRTVQCTDGS 1505

Qy 1507 SFVDEVEKVKCGCTC 1523

Db 1506 SFVEVERHLEGGCRAC 1522

RESULT 8

O75093

ID O75093 PRELIMINARY; PRT: 1534 AA.

AC O75093;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-AGO-1999 (TrEMBLrel. 11, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE SLIT-1 PROTEIN.

GN SLIT-1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99033071; PubMed=9813312;

RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;

RT "Cloning and expressions of three mammalian homologues of Drosophila

RT slit suggest possible roles for Slit in the formation and maintenance

RT of the nervous system.";

RL Brain Res. Mol. Brain Res. 62:175-186(1998).

DR EMBL: AB017167; DAB35184.1; -.

DR HSP: P00743; ICFP

DR INTERPRO: IPR000152; -.

DR INTERPRO: IPR000359; -.

DR INTERPRO: IPR000372; -.

DR INTERPRO: IPR000483; -.

DR INTERPRO: IPR000561; -.

DR INTERPRO: IPR000742; -.

DR INTERPRO: IPR001438; -.

DR INTERPRO: IPR001611; -.

DR INTERPRO: IPR001791; -.

DR INTERPRO: IPR001881; -.

DR INTERPRO: IPR002049; -.

DR PFM: PFM0008; EGF; 9.

DR PFM: PFM0004; laminin_6; 1.

DR PFM: PFM0060; LRR; 19.

DR PFM: PFM1462; LRRNT; 4.

DR PFM: PFM1463; LRCHT; 4.

DR PRINTS: PR00010; EGFBLD.

DR PRINTS: PR00011; EGFPLAMININ.

DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.

DR PROSITE: PS00022; EGF_1; UNKNOWN_3.

DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.

DR PROSITE: PS01185; EGF_2; 8.

DR PROSITE: PS01187; EGF_3; 2.

DR PROSITE: PS01225; CTCK_2; 2.

KW Glycoprotein; EGF-like domain.

SQ SEQUENCE 1534 AA; 167951 MW; 8954F8B4A8B81 CRC64;

Query Match 67.3%; Score 5597; DB 4; Length 1534;

Best Local Similarity 65.3%; Pct. No. 0;

Matches 992; Conservative 228; Mismatches 283; Indels 16; Gaps 5;

Qy 15 LVAILNVAQPAQCAQCSGSDVCHGLASVFPNIPNTERLQDNGNNITRFTED 74

Db 21 LLMAANWLAGALCAPCTCTCTTQVCHGLGAIQAIKPNIPNTERLQDNGNNITRFTED 80

Qy 75 FAGLHRLVILWENKINISIERGAFQDKLELRRLNRLNHLQPELLFLGTAKLYLRLD 134

Db 81 FAGLQRLVILWENKINISIERGAFQDKLELRRLNRLNHLQPELLFLQNNQALSRLD 140

Qy 135 SEWQIATIPRAFGAVDINKLQVYISICIDGAFRLQDLVILANNKNTIPVSS 194

Db 141 SENAIQIATIPRAFGATDKNLRLDKNYSICIEGAFRLQDLVILANNKNTIPVSS 200

Qy 195 FNMFKLRTFSLSSNNLYCDHLANLQWLKRPFGVLITQCHMPSHLGHVAVQKRE 254

Db 201 FNMFKLRTFSLSSNNLYCDHLANLQWLKRPFGVLITQCHMPSHLGHVAVQKRE 260

QY	255	FVCSDEEBZSCSFMARSCSVL--HCPACRCSANVDCRGKGLIEPTMLPMTIEIRLE	312
Db	261	FSCSGQZAGR--VPTCTLSBSCSFMARSCSVLDCRGKGLIEPTMLPMTIEIRLE	317
QY	313	QWIKIYIPGPAFSPPFKLARILDSNQISLADPAQGLRSLSNSVLGNKLTIEPMSIF	372
Db	318	LNKISYIPGPAFSPPFKLARILDSNQISLADPAQGLRSLSNSVLGNKLTIEPMSIF	377
QY	373	BGLFSQLLLLNANKINCLRWADPQGLNHLNLSLNDKLTITAGKTSPLRAITQWHLA	432
Db	378	GLGLTYQLLLLNANKINCLRWADPQGLNLSLNDKLTITAGKTSPLRAITQWHLA	437
QY	433	QNPFTCDCHLWADLTIRNPITZSGARCTSPRLNKRITQIKSKAFKCS-----G	488
Db	438	QNPFTCDCHLWADLTIRNPITZSGARCTSPRLNKRITQIKSKAFKSAKEQITPG	497
QY	485	TEDTSLKSGDCPADLACEKRCBGTVDSCNQKLNIPBIPQYTAELRANNEPTVL	544
Db	498	TEDY--QLNSCENSDVCPHCRCEANVSCSISLCTITPERIPQYTAELRANNEISL	555
QY	545	EATGIFPKLPQLRKINFSNKITDIEEGAFEGASVYELLTSNRLNVDQHMFGLES	604
Db	556	EATGKFKLTILKILNLSNKVSEDEGAFEGAASVEHLHTANQLSITSCMGFQGL	615
QY	605	KTMLRNSNRITCVGNSDIFGLSSVRLSLDNDQITTVAGAPDTLHSLISTANLANPFC	664
Db	616	KTMLRNNRISCTDIFGLSPLNLRVLSLNDQITTVAGAPDTLQSLSTANLANPFC	675
QY	665	NCTLAWLGNKRKRITVGNPACQVPLKRIPIQDAIQDPTCDGNDNCSLRCP	724
Db	676	NQCLAWLGNKRKRITVGNPACQVPLKRIPIQDAQVDFPCBEGBGGCLRPQPC	735
QY	725	TBCTCTDLYVRCNSMGLKVLKPGIPDRTVELYDGNQITVLPKLSNKRHLTSLDNNR	784
Db	736	QBACATDLYVRCNSKRRLALPGIPKMNVELYDGNQITVLPGLSLFTYQLVLSLNN	795
QY	785	ISTLSQSFNSINMTLLTSLNWRKRLPPTPTDGLSKRLSLSHGNDISVPEGANDL	844
Db	796	ISSLSNSPTNMSQLTLLSTANQCPITPLAPQGLSKRLSLSHGNDISTQEGIPADV	855
QY	845	SALSHLAGANPLDCNMQNSDWKSEYKEPGIARCAQCGDEAKMLTITPSKFTQ	904
Db	856	TSLSHLAGANPLDCCHLWNSWVKTYKEPGIARCAQCGDEKMLTITPAKFTQ	915
QY	905	GPVDVWILAKNCLPSNCKNGCTNSDPVDFRCPCPTFGKQGDQCVTHACISNPKH	964
Db	916	GPPTLVAQKCDLSSPCQNGCTNDHPLEVRKACPSYKGRDCVSLNCSGSPCEN	975
QY	965	GGTCHLKGEGDQFWICADGFEKNCVNDQDCEBNDCKNSCTDGVNINNYTCPPET	1024
Db	976	GGTCAQGEQDAPFTCSPTGEGPTCVNDQDVAHCAMGVGVGVGNTCQCPQLT	1035
QY	1025	TGELCEEKLDCAQDLNCPQHSKCLITPFGKCDCTPTGVGEBCHDFDQDQKCKNG	1084
Db	1036	EGACQGLVDLCSPLDNQHQEACVCTPTGPGRCBMPGYAGDNCSEMQDCKRHCNG	1095
QY	1085	AHCTDAVNGTICLPGEYSLGECRFSPPWMLPRTISPCDNFDQNGACQIVRINEPICQL	1144
Db	1096	AQCDHEVNSTSCBAGSEYSGQLCPIPHPLPRA--SPCESTEQNGACVQGNRCVQCL	1154
QY	1145	PGYQGEKCKLVSYFNINKESLYLIPSAKVPQNTIQLIATDEDSGILLYKGDHIAV	1204
Db	1155	PGFGPGECKLSVNFVDRTQLTQDLQWFRANTICLQVSTAEQDILLYKGDHIAV	1214
QY	1205	ELYRGEVRASTDTGSPASATISYETINDGFIHVALLDQSLSDVGGNPKITZLS	1264
Db	1215	ELYQGEVHVSDTSGSPASATISATINDGFIHVTVLAFQDMSIDGSPMINDNF	1274
QY	1265	KQSTLNFDSPLVYGGMPKSNVSLRQAPQNGTSTFHGICINLYINSELQDQKVPMTQ	1324
Db	1275	KHYTLNSAPLYVGGVVDVNSAARLQWLQDQSTFHGICINLYINSELQDQKTPWQ	1334

[illegible]

```

0882779
ID 088279 PRELIMINARY; PRT; 1531 AA.
AC 088279;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MEGF4.
GN MEGF4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=96360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Momura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RL like motifs by motif-trap screening.";
RL Genomics 51:27-34 (1998).
DR EXBL: AB011530; BAA32460.1; -.
DR HSP: P00743; 1CCF.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000359; -.
DR INTERPRO: IPR000372; -.
DR INTERPRO: IPR000483; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000742; -.
DR INTERPRO: IPR001438; -.
DR INTERPRO: IPR001611; -.
DR INTERPRO: IPR001791; -.
DR INTERPRO: IPR001881; -.
DR PFAM: PF00008; EGF_9.
DR PFAM: PF00054; laminin_G; 1.
DR PFAM: PF00560; LRR; 19.
DR PFAM: PF01462; LRRT; 4.
DR PFAM: PF01463; LRRT; 4.
DR PRINTS: PRO0010; EGFBL00D.
DR PROSITE: PS00010; ASX_HDXTXL; UNKNOWN_2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 8.
DR PROSITE: PS01187; EGF_CA; 2.
DR PROSITE: PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1531 AA; 167497 MW; DPC460CC8C5529A CRC64.

```

Query Match 67.14; Score 5578.5; DB 11; Length 1531;
Best Local Similarity 65.04; Pred. No. 0;
Matches 988; Conservative 230; Mismatches 282; Indels 19; Gaps 6;

QY 15 LVAILNKVAPQCACPAQCSGSGTVDCHGLALRSVPNIPRINTERLDLGNNTITKTD 74
 | | :
DB 21 LVAAGAKIKCCDPAICCTCTGVVCHGTGIAIPNPPIPPFIFNGNNITPHEKD 80

Qy 75 FAGLRHLRVQLMKNKISTIERGAFDQKLELERLRLNRNHLQFPELLFLGTAKLYRDL 134
 Db 81 FAGLRHLRVQLMKNQIGAVERRAGDQKLELERLRLNRNHLQFPELLFQNNQALSRLD 140

Qy 135 SENQIQAPKPAFAGVADIKNLQDYNQISIEGAFRALDLEVLNNNNITRLSVA 194
 Db 141 SENSLQAPKPAFAGVADIKNLQDYNQISIEGAFRALDLEVLNNNNITRIPVS 200

Qy 195 FNMFKPRTFLRSLNNLYCDHLAWLSOWLKRPRVGLTQTCQPSLRHGNVAVQRE 254
 Db 201 FNMFKPRTFLRSLNNLYCDHLAWLSOWLKRPRVGLTQTCQPSLRHGNVAVQRE 260

Qy 255 FVCSDEEHHQSPMAKCSVL--HPCAACSNVIVDCRGKLTETPILNPETIEIRLE 312
 Db 261 FVCSGQEAQAQ--VPACTLSSGSCAMKCSGIVDCRGKLTETPILNPETIEIRLE 317

Qy 313 QNTIKVTPGAFSPYKLRRLDLSNQISLAPDAGFGLSLNLVLYGNKITLPKSL 372
 Db 318 LNCIKVTPGAFSPYKLRRLDLSNQISLAPDAGFGLSLNLVLYGNKITLPKSL 377

Qy 373 EGLFSLQLLNNANKINCLVDFADQLNNLNLISLNYKLIQIAGTFSRLAQTMLA 432
 Db 378 EGLTFLQLLNNANKINCLVDFADQLNNLNLISLNYKLIQIAGTFSRLAQTMLA 437

Qy 433 QNPFICDCHLWADYLTNPETSGARCTSPRLANKRIGQIKSKFPCS-----G 484
 Db 438 QNPFICDCHLWADYLTNPETSGARCTSPRLANKRIGQIKSKFPCS-----G 497

Qy 485 TETYSKLSGDCFDALACPKRCBEGTVDCSNQKLNKPEHPIQYTAELNNKEFTVL 544
 Db 498 TETHD--LNSBCTSDVACPKRCBESVDCSNQKLSKPERIPQSTELNNKEISL 555

Qy 545 EATGIFKKLPQKLRNFSNKKITDEEGAFGASGVNELLNKRLENVQHMKYGLSL 604
 Db 556 EATGLFKKLSHLKRLNLSNKKVSEIEDQTPGATVSELHFMGLSVSGMSGLDGL 615

Qy 605 KILRLSNRITCVNDSPFGLSVELLSLDNQITVAPGATDHLRSLTLNLANPNC 664
 Db 616 KILRLSNRITCVNDSPFGLNVLRLSLDNHTTISPGATDQLSLTLNLANPNC 675

Qy 665 MYLNLWGLWKKRRIYVGNPQAPFLKPEIPIQVADIQDTCOGNDNCSPLSCP 724
 Db 676 MYLNLWGLWKKRRIYVGNPQAPFLKPEIPIQVADIQDTCOGNDNCSPLSCP 735

Qy 725 TECTCTDVTWCSNKLGLVLPKGIPOVTELYLDGNQVLPKZLSNKKHLILDSNNR 784
 Db 736 TECTCTDVTWCSNKLGLVLPKGIPOVTELYLDGNQVLPKZLSNKKHLILDSNNR 795

Qy 785 ISTLSNQSFSNMVLTLLLSNLRKIPPTDGLSLRLSLHGNDSVYVPEFADL 844
 Db 796 ISTLSNQSFSNMVLTLLLSNLRKIPPTDGLSLRLSLHGNDSVYVPEFADL 855

Qy 845 SALSHLAIAGNPLICDQNNQWSDWVSEYKEPGIARCAAGEMAKLLTTPSKFTQ 904
 Db 856 TSLSHLAIAGNPLICDQNNQWSDWVSEYKEPGIARCAAGEMAKLLTTPSKFTQ 915

Qy 905 GPVDVNLAKNCPCLSNPCNDGTNSDPVDFRTCTCPYGGQDCVPIHACISNCKH 964
 Db 916 GPVSLAVAKNCPCLSNPCNDGTNSDPVDFRTCTCPYGGQDCVPIHACISNCKH 975

Qy 965 GTGCHKEBEGDFWICADGFEKNCVNVDDCEDNENNSCTVDGINNYTCLCPPEY 1024
 Db 976 GTGCHKEBEGDFWICADGFEKNCVNVDDCEDNENNSCTVDGINNYTCLCPPEY 1035

Qy 1025 TGELCEKDLFQADLPACQHDHDKSLTLPKFGKDCDCTPGYVGEHCDIFDQDQNKRC 1084
 Db 1036 TGACRQVDFQADLPACQHDHDKSLTLPKFGKDCDCTPGYVGEHCDIFDQDQNKRC 1095

Qy 1085 AHCTDAGVNYGTCICPEGYSLGFCEFSPPMVLPSNCDNFCQNGAICVIRNEPICQL 1144
 Db 1096 ACQVDEINSTACLCAGYSGGLCEIPP---AFRNS--CBETEQNGANCVDQSSRPVQCL 1151

Qy 1145 PGTCGCEKXKLVSVNFVNKESLTIQPSAKVQPTNITLQIATDEDSGILLYGDKDAIV 1204
 Db 1152 PGCGPCEKXKLVSVNFVDDTLTQDLPQNPWRITLQDSTADNDSGILLYGDKDAIV 1211

Qy 1205 ELYRGEVASTDGTGSHPASAIYSVETINDGNHIVELLALDQSLSDVGGNPIITLNS 1264
 Db 1212 ELYRGEVASTDGTGSHPASAIYSVETINDGNHIVELLALDQSLSDVGGNPIITLNS 1271

Qy 1265 KQSTLNDPSLYVGMGPKSNVASLRQAPQNGTSHFCICNRLNINSELQDFQKVMQTG 1324
 Db 1272 KHTLNSAPLYVGMGPKSNVASLRQAPQNGTSHFCICNRLNINSELQDFQKVMQTG 1331

Qy 1325 ILPCGCEKXKLVSVNFVNKESLTIQPSAKVQPTNITLQIATDEDSGILLYGDKDAIV 1384
 Db 1332 VPGCGPCEKXKLVSVNFVDDTLTQDLPQNPWRITLQDSTADNDSGILLYGDKDAIV 1391

Qy 1385 NAFSTCKLQGHGVLCDREEDLNFQCAIKCHKGKRLSGGLQPYPCSSGTYDSCD 1444
 Db 1392 DALAYSCQDQDQSGALCMQVGAERPQGLQCLHQAQASATGACVYSPGFSGLCE 1451

Qy 1445 REISCRGERIDYRQKQYAACTTKVSLRCLRGCGAGGAGGCGPCLRRKRYSEFTCD 1504
 Db 1452 QESCRGERIDYRQKQYAACTTKVSLRCLRGCGAGGAGGCGPCLRRKRYSEFTCD 1511

Qy 1505 GSSFVDEKVKVQCTRC 1523
 Db 1512 GTSFAVEKPYKQCAPC 1530

RESULT 10
 Q94WB5
 ID Q94WB5 PRELIMINARY; PRT: 1531 AA.
 AC Q94WB5;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SLITI.
 GN SLITI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER/ICR;
 RA Yuan W., Zhou L., Chen J.-H., Wu J.Y., Rao Y., Ornitz D.M.;
 RT "The mouse SLIT family: Secreted ligands for Robo expressed in
 patterns that suggest a role in morphogenesis and axon guidance."
 RL Dev. Biol. 0:0-0(1999).
 DR ENBL: AF144627; AAD4758.1; -.
 DR HSDP: P00743; 1CCP.
 DR MDG: MGI:1315203; Slit1.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000359; -.
 DR INTERPRO: IPR000372; -.
 DR INTERPRO: IPR000483; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR000742; -.
 DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001611; -.
 DR INTERPRO: IPR001791; -.
 DR INTERPRO: IPR001881; -.
 DR PFAM: PF00008; BGF; 9.
 DR PFAM: PF00054; lamininG; 1.
 DR PFAM: PF00560; LRR; 19.
 DR PFAM: PF01462; LRRT; 4.
 DR PFAM: PF01463; LRRT; 2.
 DR PRINTS: PR00010; BGFBL00.
 DR PROSITE: PS00010; ASL_HYDROXYL; UNKNOWN.2.
 DR PROSITE: PS00022; BGF.1; UNKNOWN.9.
 DR PROSITE: PS01185; CTCK.1; UNKNOWN.1.
 DR PROSITE: PS01186; BGF.2; 8.
 DR PROSITE: PS01187; BGF_CA; 2.

SQ SEQUENCE 1531 AA; 167545 MW; F7D09AA6693A4F30 CRC64;

Qy 905 GPVDVNWILAKNCPLSNPCKNDGTNSDPVDFYRCTCPGYFGQDCDVPVHACISNPCKH 964
 || : : |||:||||:|:| |||::|:: ||||| |:|: |:| : | ||| :
 Db 916 GPPSLAWQAKDCPLSSPCNGTCHNDPLEVYRCTCPSGYKGRHCEVSLDGSSNPGCN 975

Db 1512 GTSFAEEVEKPTKCGCAQCQV 1531

DR INTERPRO; IPR000742; -.

DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001611; -.
 DR INTERPRO: IPR001791; -.
 DR INTERPRO: IPR001881; -.
 DR INTERPRO: IPR002049; -.
 DR PFAM: PF00008; EGF; 9.
 DR PFAM: PF00054; laminin_G; 1.
 DR PFAM: PF00560; LRR; 7.
 DR PFAM: PF01462; LRRNT; 2.
 DR PFAM: PF01463; LRRCT; 2.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR00011; EGFFLAMININ.
 DR PRINTS: PR00019; LEURICHP.
 DR PRINTS: PR00287; THIONIN.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 7.
 DR PROSITE: PS01187; EGF_CA; 2.
 DR PROSITE: PS01225; CTCK_2; 1.
 KW Glycoprotein; EGF-like domain.
 FT NON_TER 1 1
 SQ SEQUENCE 1025 AA; 112974 MW; 46CD0D5B7246C72 CRC64;

Query Match 66.54; Score 5530; DB 11; Length 1025;
 Best Local Similarity 95.94; Pred. No. 0;
 Matches 983; Conservative 25; Mismatches 17; Indels 0; Gaps 0;

Qy 501 ACPEKCRCEGTIVDCSNQKLNKIPHIPTTAELRNNNEPTVLEATGIFKLPQLRKIN 560
 Db 1 ACPEKCRCEGTIVDCSNQKLNKIPHIPTTAELRNNNEPTVLEATGIFKLPQLRKIN 60

Qy 561 FSNKNTLIEEGAFEGASGYVEILLTSNRLNWHQKMGKLSLSTLMLSNRITCVGND 620
 Db 61 FSNKNTLIEEGAFEGASGYVEILLTSNRLNWHQKMGKLSLSTLMLSNRITCVGND 120

Qy 621 SFVGLSVKLLSLDNQITTVAPAGFDLHSLSTLMLANFNPCNCTLAWGLWLAAR 680
 Db 121 SFVGLSVKLLSLDNQITTVAPAGFDLHSLSTLMLANFNPCNCHLAWGLWLAAR 180

Qy 681 VTGNPQACQPTFLKEIPQDVAIQDFTCDDGNDNSCSPLSKPSECTCLITVRCNSNG 740
 Db 181 VTGNPQACQPTFLKEIPQDVAIQDFTCDDGNDNSCSPLSKPSECTCLITVRCNSNG 240

Qy 741 LAVLPKGIKPDVTELYLDGNQPLVPKLSNWKHLTLDLSNNRISTLSNQSFNNQQL 800
 Db 241 LAVLPKGIKPDVTELYLDGNQPLVPKLSNWKHLTLDLSNNRISTLSNQSFNNQQL 300

Qy 801 TLLSTNRLCIPPTFDGLSLRLLSLHGNDISVYPGAFNDLSALSHLAGANPLYCD 860
 Db 301 TLLSTNRLCIPPTFDGLSLRLLSLHGNDISVYPGAFNDLSALSHLAGANPLYCD 360

Qy 861 CNQNLSDWKVSEYKEPGIARCAAGGEMADKLLLTSPKKTFCQGPMDITIAQMCPLS 920
 Db 361 CNQNLSDWKVSEYKEPGIARCAAGGEMADKLLLTSPKKTFCQGPMDITIAQMCPLS 420

Qy 921 NPKENDGTCSNDPVDYFRTCTPYGKSGDCDVPFHACISNPKHGTCHLKEBEGDPC 980
 Db 421 NPKENDGTCSNDPVDYFRTCTPYGKSGDCDVPFHACISNPKHGTCHLKEBEGDPC 480

Qy 981 ICADGPEBGENCWNVDCECDNENNSTCVDGINNTYCLCPETTGELCEKLDPCAQL 1040
 Db 481 ICADGPEBGENCWNVDCECDNENNSTCVDGINNTYCLCPETTGELCEKLDPCAQL 540

Qy 1041 NQCQDSKCLILPKFGKCDCTPGYVGEHCDIDPDCDNKCKNGAHCDAVNGYTCICPE 1100
 Db 541 NQCQDSKCLILPKFGKCDCTPGYVGEHCDIDPDCDNKCKNGAHCDAVNGYTCICPE 600

Qy 1101 GYSLGCFSPPMVLPRTSPCDNFDCNGAGCIIRINEPQCCLPGYQKCEKLSVNF 1160
 Db 601 GYSLGCFSPPMVLPRTSPCDNFDCNGAGCIIRINEPQCCLPGYQKCEKLSVNF 660

Qy 1161 INKSTLIQIPSAKVRPQTNIITLIATDEDSGILLYGDKDHIAVLYGRVASYDTGSH 1220

Db 661 INKSTLIQIPSAKVRPQTNIITLIATDEDSGILLYGDKDHIAVLYGRVASYDTGSH 720

Qy 1221 PASAITSYVETIDNCFHIVELLALQDLSLSVDGGNPKITNLKSQSTANFDSLTVGGM 1280

Db 721 PASAITSYVETIDNCFHIVELLTDLSSLSLSVDGGSFKVITNLKSQSTANFDSLTVGGM 780

Qy 1281 PGKNVSAIRLQAPQNGTSGFHCIRNLINSELQDPQKVMPTGILPGCEPCCHKVCAG 1340

Db 781 PGKNVSAIRLQAPQNGTSGFHCIRNLINSELQDFRKMPMTGILPGCEPCCHKVCAG 840

Qy 1341 TCQPSQAQGTCTCQEGWMPGLCDQRTNDPLCNKCVHGTCLPINFATSCKLEBHG 1400

Db 841 TCQPSQAQGTCTCQEGWMPGLCDQRTNDPLCNKCVHGTCLPINFATSCKLEBHG 900

Qy 1401 LCDREEDLNFQCAIKCKHGCRLSGLGOPYCFCSSGTGSDCREISCRGERIDTYK 1460

Db 901 LCDREEDLNFQCAIKCKHGCRLSGVGOPYCFCNSGTFGSDCREISCRGERIDTYK 960

Qy 1461 QQTVAACQITKVKSLRLECGGAGGCGPLSKRKRKYSPFCTDSSSFVDEYKVKCC 1520

Db 961 QQTVAACQITKVKSLRLECGGAGGCGQCPKSKRKRKYSPFCTDSSSFVDEYKVKCC 1020

Qy 1521 TRCVS 1525

Db 1021 ARCAS 1025

RESULT 12
 Q9WU5
 ID Q9WU5 PRELIMINARY; PRT; 1530 AA.
 AC Q9WU5
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SLT12.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99200391; PubMed=10102268;
 RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
 RA Tessier-Lavigne M., Kidd T.;
 RT "Slit proteins bind Robo receptors and have an evolutionary
 RT conserved role in repulsive axon guidance."
 RL Cell 96:795-806(1999).
 DR EMBL: AF133730; AAC2540.1; -.
 DR HSSP: P00743; 1CCF.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000359; -.
 DR INTERPRO: IPR000372; -.
 DR INTERPRO: IPR000483; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR000742; -.
 DR INTERPRO: IPR001438; -.
 DR INTERPRO: IPR001611; -.
 DR INTERPRO: IPR001791; -.
 DR INTERPRO: IPR001881; -.
 DR INTERPRO: IPR002049; -.
 DR PFAM: PF00008; EGF; 9.
 DR PFAM: PF00054; laminin_G; 1.
 DR PFAM: PF00560; LRR; 18.
 DR PFAM: PF01462; LRRNT; 3.
 DR PFAM: PF01463; LRRCT; 4.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR00011; EGFFLAMININ.
 DR PRINTS: PR00019; LEURICHP.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.

DR PRINTS; PRO1143; FSHRRECEPTOR.

SQ SEQUENCE 796 AA; 89542 MW; 0F50806FC0345005 CRC64;

Query Match 47.3%; Score 3934; DB 11; Length 796;
 Best Local Similarity 96.54%; Pred. No. 0;
 Matches 740; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MRGVGQWMLSLGLVLAAILNKVAPQACPAQCSGSGVCHGLALRVSVPNIPNTERLE 60

Db 1 MCGISQWMLSLGLVLAAILNKVAPQACPAQCSGSGVCHGLALRVSVPNIPNTERLE 60

Qy 61 DLNGNNTITRTITPAGLRHLRLIQLMENKSTIERGAFDLKLERLKNRNQLQFPE 120

Db 61 DLNGNNTITRTITPAGLRHLRLIQLMENKSTIERGAFDLKLERLKNRNQLQFPE 120

Qy 121 LLFGTAKLRLQSENQIQAPKFAFGAVDINKLQDYNQISCEIDGAFRALRDLVLE 180

Db 121 LLFGTAKLRLQSENQIQAPKFAFGAVDINKLQDYNQISCEIDGAFRALRDLVLE 180

Qy 181 TLNNNITRLSVAFSNMPLKRTPLHLSNNLCOCHLANSQDLKRRPVGGLTQCMGSP 240

Db 181 TLNNNITRLSVAFSNMPLKRTPLHLSNNLCOCHLANSQDLKRRPVGGLTQCMGSP 240

Qy 241 HLRGHNVAVQKRFVCSDEEBGHQSFMAPSCSVLHPACITCSNNIVCGRGLTETP 300

Db 241 HLRGHNVAVQKRFVCSDEEBGHQSFMAPSCSVLHPACITCSNNIVCGRGLTETP 300

Qy 301 NLPTITLRLQSNITVIPPQAFSPYKLRRLDNNQISLAPQGLRSLNSLVLY 360

Db 301 NLPTITLRLQSNITVIPPQAFSPYKLRRLDNNQISLAPQGLRSLNSLVLY 360

Qy 361 GKKITLPSLPSGLSLQLLLNANKINCLRVAPQDLHNLNLSYDNKLTQVAGT 420

Db 361 GKKITLPSLPSGLSLQLLLNANKINCLRVAPQDLHNLNLSYDNKLTQVAGT 420

Qy 421 SPRLAQTMLHQAQNPFCOCHLAWADYLNHNISGARTCSPRLANRIGQIKSKF 480

Db 421 SPRLAQTMLHQAQNPFCOCHLAWADYLNHNISGARTCSPRLANRIGQIKSKF 480

Qy 481 RCGSTGYTSKSLGDCFDLACPKRCRCBGTVCSNQKLNKIPHIQPTAELRANNE 540

Db 481 RCGSTGYTSKSLGDCFDLACPKRCRCBGTVCSNQKLNKIPHIQPTAELRANNE 540

Qy 541 FTVLEAFGPKLPQLARKINFSNNKITIEBGAFGASGVNELLTSNRLNVQHMFGP 600

Db 541 FTVLEAFGPKLPQLARKINFSNNKITIEBGAFGASGVNELLTSNRLNVQHMFGP 600

Qy 601 LESLKTMLASNRITCVGNDSFTGLSSVRLSLYDNNITVAPGFTLASLSTANLAN 660

Db 601 LESLKTMLASNRITCVGNDSFTGLSSVRLSLYDNNITVAPGFTLASLSTANLAN 660

Qy 661 PFNCCNLAHLEWLRKRRTVGNPCQKPYFLKEPIQVADQPTCDGNDGNSCPL 720

Db 661 PFNCCNLAHLEWLRKRRTVGNPCQKPYFLKEPIQVADQPTCDGNDGNSCPL 720

Qy 721 SRCPTCTCLTVVRCSSNGLKVLPGKIPDVTLEYLDGNQFLVLPK 767

Db 721 SRCPTCTCLTVVRCSSNGLKVLPGKIPDVTLEYLDGNQFLVLPK 767

RESULT 14

Q9V7F9

ID Q9V7F9 PRELIMINARY; PRT; 1504 AA.

AC Q9V7F9;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE SLI PROTEIN.

GN SLI.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

EX MEDLINE=20196005; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chanpe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Belt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.P., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Wray A.D., Dew I., Dietz S.M.,

RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin J.K., Evangelista C.C., Ferraz C., Ferrieria S., Fleischman J.,

RA Fostler C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,

RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris M.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin P., Houston K.A., Howland T.J., Wei M.-H., Iyegyan C.,

RA Jalali M., Kalush F., Karpov G., Ke Z., Kennison J.A., Ketchum K.A.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattet B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Munny D.M., Nelson D.L.,

RA Nelson D.M., Nelson K.A., Nixon K., Nusskera D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders A.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-T., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.C.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.

RT "The genome sequence of Drosophila melanogaster."

RL Sequence 287:2185-2195(2000).

DR EMBL: AB003809; AAF58097.1; -.

DR HSSP: P00743; 1CCF.

DR FLYBASE: FBgn0003425; sli.

DR INTERPRO: IPRO00152; -.

DR INTERPRO: IPRO00359; -.

DR INTERPRO: IPRO00372; -.

DR INTERPRO: IPRO00483; -.

DR INTERPRO: IPRO00561; -.

DR INTERPRO: IPRO00742; -.

DR INTERPRO: IPRO01611; -.

DR INTERPRO: IPRO01791; -.

DR INTERPRO: IPRO01881; -.

DR INTERPRO: IPRO02049; -.

DR PFAM: PF000007; Cys_knot.1.

DR PFAM: PF000008; BGF.7.

DR PFAM: PF000054; laminin_G.1.

DR PFAM: PF000560; LRR.17.

DR PFAM: PF01462; LRRNT.4.

DR PFAM: PF01463; LRRCT.4.

DR PRINTS: PR00011; EGF_LAMININ.

DR PRINTS: PR00019; LEURICHRPT.

DR PROSITE: PS00010; ASX_HYDROXYL.3.

DR PROSITE: PS00022; EGF_1.7.

DR PROSITE: PS01185; CTCK_1.1.

DR PROSITE: PS01186; EGF_2.5.

DR PROSITE: PS01187; EGF_CA.2.

DR PROSITE: PS01225; CTCK_2.1.

SQ SEQUENCE 1504 AA; 168597 MW; 836A3F5022BF234F CRC64;

Query Match 43.1%; Score 3588; Db 5; Length 1504;
 Best Local Similarity 44.9%; Pred. No. 2.6e-277;
 Matches 677; Conservative 273; Mismatches 466; Indels 92; Gaps 21;

Qy 28 CPAQCSGSGTVDCHGLALRSVPNPRINTERLIDNGNNTIRITKIDFAGLRHLVLQML 87
 Db 73 CPWFVSCGTLGNDCHSHGLSVPRKISADVERLGLQGNVITVYDTPQRLTLRLQLUT 132
 Qy 88 ENKISTIERGAPQDLKELERLNRNLHQLPELFLGTAKYRLDLSNQIQAIPKAF 147
 Db 133 DQNIHTIERNSPQDLVSLERLNNRKLKAIPEFNVTSSASLRDIDLSNVITTVGRVY 192
 Qy 148 RGAVIDNKLQDYNQISCIEDGAFRALDLEVLITNNNNITLSVAFSNMPLKRTFLH 207
 Db 193 RGAQSLRSQIDNNQITCLDERAFGLVLELITLNNNNITSLPHNIPGGLRLRLALS 252
 Qy 208 SNKLYCDCHLAWLSDWLKRFPVGLYTCQMGPSLRGHVNAVQKREYFVDEEHHQSF 267
 Db 253 DNPFAFCDHLSWLSRFLSATRLAPTYACQSPQLAQWADHQQEFCSGLTE----- 307
 Qy 268 MAP-SCSYLV-CPAQCSTNNIVDCRGKGLTEPTNPEITTELRQNTIKVPPGAFS 325
 Db 308 HAPMCEGANSCHPCRCAGDVCDEKRLSVTPVTLPDOTLELRQNTIFTELPKPSF 367
 Qy 326 PYKKLRIDLSNQISLAPADQGLRSNLVLYGNKITEPKSLFELFSLQLLLANA 385
 Db 368 SFBRARIDLSNNKISITADLSLQKLTUTLYLYGNKIKDLPSPGVKGLSLQLLLANA 427
 Qy 386 NKILKLRIDFADQHLNLLSLYKLTQAKGFSPLAIDTQWGLNPFICDCKRLWL 445
 Db 428 NEISCIKDPADFLSLSLSDNNQISLANGTIDAMKSITVHLANKPFCIDDLRLWL 487
 Qy 446 ADTLNPNITSGARCTSPRLKLRIGQIKSKFCSGTEDYKSLSDCFADALCEK 505
 Db 488 ADTLNPNITSGARCTSPKMRHRRISLREKFKCS-WDELRLKLSGCRNDSDCAP 546
 Qy 506 CRGEGTVDGNSQNLKIPHEIPYTAELRANNEFTVLAETGFKKLPQLRKNSNNK 565
 Db 547 CRGEGTVDGTCGKLEIPROPLATTELLDNLDELGRISDGLGRPLVLELKRQ 606
 Qy 566 ITDIEBEGAFGASGVNELLTSNRLNVOHMFKGLSLKTLMLRSNRTICVGNDSFGL 625
 Db 607 LTGTEPNAFEGASHIQLQENKIKELSNKMFGLRLQKLT----- 647
 Qy 626 SSVRLSLSDNQITVAPGADTDLSSLANLAFNFCNCLANLWGLAKKRIRVITGNP 685
 Db 648 ----LNLNDQISCVMPGSEHLNLSLMLANSPFNCHLALNFAWMLKRLKSLGAA 702
 Qy 686 RCKPFLFKETPIDQVADFTCDGNDNSCSPLRCPTCTCTIVVRCNSKGLKVP 745
 Db 703 RCGAPSKVQWIKDLPSEFKCSSENS-GLGDGTCFPGSCITGTIVVRCNSKLEIP 761
 Qy 746 KGIPDVTVELYDGNQFTLVKPE-LSNYKHLTIDLSNNRISTNSQFSNNQTLTIL 804
 Db 762 RLPARTSELFTESNEIQHYERIRHLSLRLDLSNNQITLSNTPFANLKLTLII 821
 Qy 805 SYNRKLCPPIPTFDGLSLRLSLAGNDSIVVPEGAFNDLSLHAIGANPLICDQNG 864
 Db 822 SYNLQCLQRLHSLGLANLVLISLGRNLSMLPGSPEDLKLSTHIALGNSPLTCOGSL 881
 Qy 865 WLSQWYKSEKPIEGARCAAGCGEMAKLLLTTPSKKFTQCGPVDVNLKANPCLSNPK 924
 Db 882 WFSWIKLIDVPEGIARCAPEQMCKLLLTSPSSSFCVGRVNDLILKANCFBPQ 941
 Qy 925 NQDCTNSVDVDFRCTCPYFGKQDQVPIHACISNCPKRGHTCHLKEGEDGFWICAD 984
 Db 942 NQACVQLPQREYQCLQCPYHGKHCPEMIDACYNCPNNACTVL--EBGFSQCAP 999
 Qy 985 GFEGENECVNVDC-EDNDCNNSTVDGINNYTCLCPETTEGCEERLQCAQDLNCP 1043
 Db 1000 GYTGARCTNIDCLGEIKQNNATCIDGVESYKCRQCPGSGEPCKTIQCPSPFNPC 1059

Qy 1044 QHSDKILPKFGKCDCTPGYGEHCDIDFDCQNKCKNGAHCATDAVNGYTCICPEYS 1103
 Db 1060 ANGAKMDHFTYHSCDQAGFHTGTCTNDIDCNHWCQNGTCDVENDYTCQPDPT 1119
 Qy 1104 GLPCEFSF-PWLPRTSPCDNFDCQNGA-QCIVRINEPIQCLPGYGEHCKELVSVN 1159
 Db 1120 GKICEGNNHISMTPTSPQNEHCHEKVCQPNAGQSDYLCRCPGTGKWEYLTSSIS 1179
 Qy 1160 FIMKSELIQPSAKVRPQNTITLQIATEDSGILLYKGDHAYHGLRVASRYDTGS 1219
 Db 1180 FVNNKSEVLEPLATPEANVTIVSSAQNGILMYDQDAILAYELWGRIRVSYDVGN 1239
 Qy 1220 HPASIVSVETINDGNFHVILLADQSLSLSDVGNPKITINLSKQSTLNFDSPLTVGG 1279
 Db 1240 HPVSTMYSEFVADKGYHAYELLAKNNTLRVDRGLARSILNEGSDYLLKLTTPMFLG 1299
 Qy 1280 MPKSNVSLAQAPQNGTSGFHCIRNLYINSELQDQKVPMTQILPGCEPCCHKVCAH 1339
 Db 1300 LPVDPQAQYKXQIKRLTSFKGCKEVIWNNKLVDFGNQKQKQITPGC-----ALLE 1353
 Qy 1340 QTCQSSQAGTCTCQBNMG-PLCDQRTNDCPLCNCHVBT-CLP-IMA-FYSKRL 1394
 Db 1354 GEQEE-----DDEQFMDETHAIKEEPVDCLENCRGRSVFNMRQGTQCKR 1407
 Qy 1395 EBGSGVLCDEEDLWNPQAIKCHKGRSLSGQYPCDCSSGYTDCSDRLSCGRRI 1454
 Db 1408 EBGSGVTCQCGSBTEP-----PVTFAAS-----TCMKQV 1438
 Qy 1455 RDTYQVQYAAQCTTKVSRLECRGCGACQCGPLRKRKRYSFECTDSSSFVDEYK 1514
 Db 1439 RYTTEND----CNRQPLKAYKVGCG-GNQCAAKTVRRKRVKNSNNRYIKNDI 1493
 Qy 1515 WKRCQCT 1522
 Db 1494 WKRCQCT 1501

RESULT 15

Q5XIV4

Q5XIV4 PRELIMINARY; PRT; 1504 AA.
 AC Q5XIV4;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SLIT PROTEIN.
 GN SLIT.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Phlebotominae; Diptera; Phlebotominae; Phlebotominae.
 RN NCBI_TaxID=7227;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93200390; PubMed=10102267;
 RA Kidd T., Bland K.S., Goodman C.S.;
 RT "Slit is the midline repellent for the robo receptor in Drosophila.";
 RL Cell 96:785-794 (1999).
 DR EMBL; AF126540; AAD26567.1; .
 DR HSP; P00740; 12EM.
 DR INTERPRO; IPR000152; .
 DR INTERPRO; IPR000359; .
 DR INTERPRO; IPR000372; .
 DR INTERPRO; IPR000483; .
 DR INTERPRO; IPR000561; .
 DR INTERPRO; IPR000742; .
 DR INTERPRO; IPR001611; .
 DR INTERPRO; IPR001791; .
 DR INTERPRO; IPR001881; .
 DR INTERPRO; IPR002049; .
 DR PFAM; PF00007; Cys_knot.1.
 DR PFAM; PF00008; DGF.7.
 DR PFAM; PF00054; laminin_G.1.
 DR PFAM; PF01462; LRRNT.4.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: January 22, 2001, 12:15:50 ; Search time 233.01 Seconds
(without alignments)
204.714 Million cell updates/sec

Title: US-09-540-245a-15
Perfect score: 7427
Sequence: 1 MHPMPENHAIARSTSTNN.....SCLYARAGEAPRQMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 268485 segs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7427	100.0	1395	20 Y13563	Drosophila Robo 1
2	7427	100.0	1395	20 Y08401	Drosophila sp. ROB
3	1788	24.1	1380	20 Y08402	Drosophila sp. ROB
4	1786.5	24.1	1381	20 Y13564	Drosophila Robo 2
5	1592	21.4	1651	20 Y13566	Human Robo 1 polyp
6	1588	21.4	1297	20 Y13565	C. elegans Robo po
7	1588	21.4	1297	20 Y08403	C. elegans ROBO pr
8	1584	21.3	1649	20 Y08404	Human ROBO1 protei
9	1317.5	17.7	753	20 W83927	Human T85 protein.
10	665.5	9.0	1571	19 W42087	Human Down syndrom
11	661	8.9	1910	19 W42086	Human Down syndrom
12	639	8.6	1728	12 R13144	Deleted in Colorec

13	637.5	8.6	1447	16 R68553	Deleted in colorec
14	637.5	8.6	1447	20 Y33498	Human DCC protein.
15	633.5	8.5	1257	20 W74152	Human Li cell adhe
16	614.5	8.3	1018	15 R63759	Human contactin (F
17	614.5	8.3	1018	17 R87028	Human contactin
18	610.5	8.2	1192	19 W57900	Protein of clone C
19	603	8.1	1299	21 Y40439	Human Nr-CAM prote
20	602	8.1	1028	19 W29667	Homo sapiens DL185
21	588	7.9	1018	18 W06485	Rat contactin like
22	586.5	7.9	1304	19 W59394	Human neural cell
23	579.5	7.8	1911	16 R71726	Human PTP-OB. Hak
24	579.5	7.8	1911	18 W27225	Human protein tyro
25	579.5	7.8	1911	20 W94027	Human protein tyro
26	573	7.7	1496	20 W81030	Melanoma associat
27	573	7.7	1496	21 Y70469	Human p53 target m
28	548	7.4	4412	21 Y33666	Sequence gi/101742
29	546	7.4	1897	21 Y81785	Human protein tyro
30	546	7.4	1897	21 Y56100	LAR tyrosine phosph
31	545	7.3	434	20 Y13567	Human Robo 2 polyp
32	545	7.3	434	20 Y08405	Human partial ROBO
33	529.5	7.1	1501	16 R72858	Rattus norvegicus
34	526	7.1	1242	19 W52287	Human PROR60 prote
35	510	6.9	985	20 Y41716	Rattus norvegicus
36	509	6.9	1225	19 W52289	Homo sapiens cdo t
37	490	6.6	3117	21 Y33667	Sequence gi/332818
38	480.5	6.5	1070	18 W08747	Human colon carcin
39	458	6.2	1853	21 Y33668	Human protein tyro
40	457	6.2	2387	21 Y33665	Mechanical stress
41	457	6.2	2597	21 Y33664	Mechanical stress
42	446	6.0	1139	19 W37779	Rattus norvegicus
43	440.5	5.9	1125	19 W52288	Rattus norvegicus
44	439	5.9	848	21 Y86565	Human NCAM 140kd 1
45	438.5	5.9	1251	19 W37778	Rattus norvegicus

ALIGNMENTS

RESULT 1
Y13563
ID Y13563 standard; Protein; 1395 AA.
XX
AC Y13563;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 1 polypeptide.
XX
RW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W09258333-Al.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 9705-0065543.
XX
PA (RBGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
XX
DR N-PSDB; X55767.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 30-33; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.

XX
 SQ Sequence 1395 AA;

Query Match 100.04; Score 7427; DB 20; Length 1395;
 Best Local Similarity 100.04; Pred. No. 0;
 Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHPMPENHAIARSTSTNNPSSRSRSMWLLPAWLLLVASNGLPANRQQTQSPRIE 60
 DB 1 mhpmpenhaiarststnnpssrsrsmwllpawlllvasnglpavrgyqsprie 60
 61 HPTDLVKNKEPATLNCKVEGKEPTIEFKDGEYVSTNEKSKSRVFKDGLAFFRTQM 120
 DB 61 hptdlvknkepatlnckvegkeptiefkdgeyvstneksksrvfkdgldaffrtqm 120
 121 GKKEGGGEGYVCKAVNRGGQAVSRHASLIQAVLRDDFVPEKDTKRAVAGETALLECGPPK 180
 DB 121 gkkegggeggyvckavnrvggqavsrhasliqlavlrddfvpekdtkravagetallecgppk 180
 181 GIPEPTLWIKDGVPLDLKANSFGASSRVRIVDGNLISNVPEIDEGNTKCIANQW 240
 DB 181 gipeptlwikdgvpldlkansfgassrvrivdgnlinsvpeidegntkciqnw 240
 241 TRESSAKLVIVQVYFKMEKQDMVLKQGTAFHCSVGDDPPKVLAKNEKINPVSR 300
 DB 241 tressaklvivqvyfmekekdmvlkqgtatfhcsvgddppkvlaknekinpvsr 300
 301 RIHDEKSLKLEISNITPTDGTGYCEAHNVGQISARASLIHVHPNFKTPSKNKVGLNG 360
 DB 301 rihdeksleisnitptdgtgyceahnvvgqisaraslihvhpntfksknkvglng 360
 361 VVGLPCMASGNPPSPVFWKEGVSTLMFPNSHKGROVVAADGLTQIDVRDEGTYVCS 420
 DB 361 vvglpcmasgnppspvfwkvgvstlmfpnshkgrovvvaadglqtqidvrdegtyvcs 420
 421 AFSVDSSTVIRVLQVSSVDERPPPIQIGPANLTKGSVATLPCRAATGSPKAFKWF 480
 DB 421 afsvdsstvirvlqvssvderpppiqigpanltpkgsvatlpcratgspkafkwf 480
 481 DGHAVAGNRSIIQSSILVLDLQLSDSGYTICTASGERGETSWAALTVEKPSGLH 540
 DB 481 dghavagnrsiiqssilvldlqlsdsgytictasgergetswaaltvekpsglh 540
 541 RAADPTYPAPGTPKVLNVSRTSISLWAKSQEKPGANGPIGITYETSPQLQGTWIV 600
 DB 541 raadptypapgtpkvlnvrsstslwaksqekpgangpigityetvspqlqgtwiv 600
 601 AAHRVDTQVTISGLTPTGTSYFLVRAENTGQISVPSGLSNVIRKTEADFAASANDLSA 660
 DB 601 aahrvdvtqvtisgltpgtstsyflvraentgqisvpsglsnvirkteadfaasandlsa 660
 661 ARTLLTGKSVELIDASAINASAVLEWMLHVSADKTYEGLRIHYKDAVSQPSQYHSITV 720
 DB 661 artlltgksvelidasainasavlewmhvsadektyegllrihykdavsqpsqyhsitv 720
 721 MDAESAESFVGNLKKYTKYEFPLTPFFETLEGQPSMSKALTIEDVSPAPDNIQIMYN 780
 DB 721 mdaesaesfvgnlkktykyefltpffetlegqpsmskaltiedvspapdniqimyn 780
 781 QTAGVWVTPPPSQHNGNLYGKYIEVSGNTMKVLNMTLNATTSTVLLNLTGAWYS 840
 DB 781 qtavwvtpppsghngnlygkyievsngntmkvlnmtlnattstvllnltgawys 840

QY 841 VRLNSFTAGUGPPYSKPSISLMDPTHVHPRAHPSGTHDGRHQGLQTLNNGNIPGD 900
 DB 841 vrlnsftagugppyskpsislmptdhvhpahpsgthdgrhggldtyhngnippgd 900
 901 INPTTHKTKTDTLSCGFWMLVLCVLLVLSIAISVYFKRKHQMTKELGHLISVDSNE 960
 DB 901 inptthktdtlysgfwmlvclvllvlsiaismvyfkrkhqmtkelghlsvdsne 960
 961 ITALINKSKESLWDHHRGWRTADTDKDSGLSEKLSHVNSQSNYNNSDGDTDAEVD 1020
 DB 961 italnkseslwdhhrgwrtadtdkdslesklsbvnsqsnynnsdgdtdaevd 1020
 1021 TRNLITFYNCRSKPNPTTATMTIGTSSSETCTKTTISADKSGTHSPYSDAFAGV 1080
 DB 1021 trnlitfyncrskpnpttattmtigtsssetctkttisadksgthspysdafagv 1080
 1081 PAVPVVKSNIQLQTPVEPINMSEFLPPPHEPPHPPSTGYAQSGPSSKSKSAGSGIST 1140
 DB 1081 pavpvksnilyqtpvepinmseflppphpphstgyaqsgpssksksagsgist 1140
 1141 NQSLINASIHSSSGSGFSWGPQYAVACPENYVSNPLSAVAGTGNRTQITQNH 1200
 DB 1141 nqslinasihsssgsgfswgpyavacpenyvnplsavagtgnrtqitqnh 1200
 1201 POLPATFATTTGGVAVPNPHLPFORHAASEYQAGLNAARCAQSRACNSDALATPSM 1260
 DB 1201 polpatfatTTGGVAVPNPHLPFORHAASEYQAGLNAARCAQSRACNSDALATPSM 1260
 1261 QPPPPVVPVPEGWYQVPHNSHMHPTSSNHQVQCSSECDHSRSSQSHKRLQLEHGS 1320
 DB 1261 qppppvvpvpegwyqvphnshmhptssnhqvqcssecdhsrssqshkrlqleehgs 1320
 1321 SAKQGGHRRHRAVPVQPCSENENMLAEYQGRYTSDCNCSREGDTCSCEGSLYA 1380
 DB 1321 sakqggghrrhRAVPVQPCSENENMLAEYQGRYTSDCNCSREGDTCSCEGSLYA 1380
 1381 EAGEPAPGMLAKNT 1395

RESULT 2
 Y08401
 ID Y08401 standard; Protein; 1395 AA.
 XX
 AC Y08401;
 XX
 DT 24-JUL-1999 (first entry)
 XX
 DE Drosophila sp. ROBO1 protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 cell morphology; screening assay.
 XX
 OS Drosophila sp.
 XX
 PN W0920764-AL.
 XX
 PD 25-APR-1999.
 XX
 PF 20-OCT-1998; 98WO-US22164.
 XX
 PR 14-NOV-1997; 97US-091712.
 PR 20-OCT-1997; 97US-0062921.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
 XX
 DR WPI; 1999-312615/26.
 DR N-PSDB; X57250.
 XX
 PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 45-49; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1395 AA;

Query Match 100.0%; Score 7427; DB 20; Length 1395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MHPFMENHAIARSTSTNNPSSRSRRSMWLLPALLVLSNGLPAVGGQYSPRIE 50
Db 1 mhpfmehaiarststnnpssrsrrsmwllpawllvlsnslpavrgyqysprlie 50
Qy 61 HPTDLVKKNEPATLCKVCKGPEPTIEWKGDFPVSTNEKSKHVRQKDGALFFYRTMQ 120
Db 61 hptdlvkknepatlckvckgpeptievkdgfpvstnekkshvrqkdgalffyrtnq 120
Qy 121 GKKEGGGTVCAKNRVQAVSRHASLQIALVRDDEVPKDALVAKGTALLBQGP 180
Db 121 gkkegggvtvcaknrvqavsrhaslqlavlrddfrvepkdtrvakgtallbqgppk 180
Qy 181 GIPEPTLWIKDQVPLDLKAMFGASSRVRIVDGNLLISNVEIDENYKICIAQNLWG 240
Db 181 gipeptliwkdqvpldlka mfgassrvrvdgnllisnveidegnykiclaqnlwg 240
Qy 241 TRSSYAKLIVQVQPKDQVWMLGQATPHCSVGDDPPVWKKREGNIVPSRA 300
Db 241 tressyakiwkvqpfmkepkdqvmlgqatphcsvgddppvwkkrengnivpsra 300
Qy 301 RILHDEKSLISNITPDEGTVCEAHNNGQISARSLIVHAPNFKRPSKVKGLNG 360
Db 301 rilhdeksleisnitpdegtvceahnngqisarslilhappnfkrpskvvkglng 360
Qy 361 VVQLPCMASGPPPVFWKVCYSLTFMFPNSHGQYVAADQLGITDVRQDEGTVFWS 420
Db 361 vvqlpcmasgpppvfwkvcysltfmpfnshgqyvaadqlgitdvrqdegvtfws 420
Qy 421 AFSVDSSTVRFVQSSVDERPPPIIQIPANGTLKGSVATLPCRACTGSPRIKWFH 480
Db 421 afsvdsstvrflvqssvderpppiiqipangtlkgsvatlpcratgsprikwfh 480
Qy 481 DGHVAGNHSYIIQSSSLAVDDQLQSLDSCITTCASGERGTSWAATLVKPKGSLH 540
Db 481 dghvagnhsyilqssslavddqlqslsdscitcasgergtswaatlvpkpgslh 540
Qy 541 RAADSPFTAPPKTPKVLNVRSTISLWAKSQRKPAVGPIIGTVETSPDLQTGWIV 600
Db 541 raadspftappktpkvlnvrrstislwaksqrkpvagpiglietvetspdlqtgwiv 600
Qy 601 AHRVGTQDTVITISGLTPTSTVFLVRAENTQGISVPSGLSNVKTIEADPAASANDLSA 660
Db 601 ahrvgtqdtvitisgltpstsvflvraentqgisvpsglsnvktieadpaasandlsa 660
Qy 661 ARTLLTKSVELIDASINASAVRLWMLHVSADENYVGLAHLKXASVPSAQYHSITV 720
Db 661 artlltksvelidasinasavrlwmlhvsadenyveglahlkxasvpsaqyhsitv 720
Qy 721 MDASAEFVGNLKKYTYEFLTPFFETIGQPSNKALTIEDVPSAPPDNIQIGNY 780
Db 721 mdasaeefvgnlkktytyefltpffetigqpsnkaltiedvpsappdniqigny 780
Qy 781 QTAGVWRTVPPTSGHNGNLGYKIEVSAGNTMVLNMLNATTISVLLNLTGAVTS 840
Db 781 qtavwrtvppptsgnhngnlgykievsagntmvlmlnattisvllnltgavts 840

Qy 841 VLNSFTKAGDGYSPKISLFWDPHTHVHPPRAHPSGTHGRHGQDLTHMNGNTPPGD 900
Db 841 vlnsftkagdygyspkislfndphtvhpprahpsgthgrhgqdlthmngnppgd 900
Qy 901 INPTTHKTKTLDLGGPWLWLCVILVILVISAALSMVTFKRKHQMTKELGHLVSDNE 960
Db 901 inptthkktldlsgpwlwlvclvvlvlsaalsmvtfkrkhmqmtkelghlvsdne 960
Qy 961 ITALINSKESLWIDHHRGWRTADTKDQSLSEKSLHSHVNSGSQVNSDGTITAEVD 1020
Db 961 italnkseslwidhhrgwrtadtkdsglsekshsvnsqsgvnsdgtitdaevd 1020
Qy 1021 TRNLTLFTNCRKSPDNPTPIATMTIGTSSESTCTKTTSISADKDSGTHSPYSDAFAGV 1080
Db 1021 trnlftlfnrcrksdpnptpiatmtlgtsssetctkttisadkdsghthspysdafagv 1080
Qy 1081 PAVPVVKSNIQYTPVEPINWSELPFPPHPPSPSTYGTAGSPSSSRKSKSAGSGIST 1140
Db 1081 pavpvyksniqytpvepinwsefpfpphppspsstygtagspsessrksksagsgist 1140
Qy 1141 NQSLINASHSSGGGFSAGVSPGYACPPENNVNSPLSANAGGTQMRQYITPTNQHP 1200
Db 1141 nqslinashssgggfsagvsgpyacppennvnsplsanaggtqmrqyitptnqhp 1200
Qy 1201 POLPFIATFTGGGAVPPNHLFPFQTHAASEYQAGLMAACQASACNSCALATPSFM 1260
Db 1201 polpfiatftgggavppnhlfpfqtthaaseyqaglmAACQASACNSCALATPSFM 1260
Qy 1261 QPPPVVPSVPEWQVPHNSHMPHTPSNHLVQCSSESDHSSSSQHWKRLQLEHGS 1320
Db 1261 qpppvvpsvpewqvpnhshmphtpsnhlvqcssecdhssssqhwkrlqleehgs 1320
Qy 1321 SAKRGYGHRRAPVWQPCMESNNWMLAEYERQRTSDCCNSREGDTCSGESCLIA 1380
Db 1321 sakrgyghrrapvwqpcmesennwmlaeyerqrtsdccnsregdtcsgeclia 1380
Qy 1381 EAGEPAPRMATKNT 1395
Db 1381 eagepaprmtaknt 1395

RESULT 3
Y08402
ID Y08402 standard; Protein; 1380 AA.
XX
AC Y08402;
XX
DT 24-JUL-1999 (first entry)
XX
DE *Drosophila* sp. ROBO2 extracellular domain protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
XX cell morphology; screening assay.
XX
OS *Drosophila* sp.
XX
PN W09920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1996; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REG.) UNIV CALIFORNIA.
XX
PI Goodman CS, Kida T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57251.
XX

FILE: 000000001 0/ 120 1/ 1200000000 10/ 120000000 0/ 1000 0/

Sequence 1381 AA:

00 049 - lmgxlyveqlyvyarqipuprvnmpapvrcnmpilqdstctasasasasailstckpna / 00

bb 135b gsvasers11shsgsgtssqpagnn 1380

RESULT 5

XX

PD 27-MAY-1999.

FF 13-NOV-1998; 98WO-0524327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RBOC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-138008/28.
DR N-PSDS; X55770.
XX
FT Modulation of Robo-Comm polypeptide Interactions
XX
XX Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commisureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
XX function.
XX
SQ Sequence 1651 AA;

Query Match 21.44; Score 1592; DB 20; Length 1651;
Best Local Similarity 30.24; Pred. Md. 3.1e-83;
Matches 419; Conservative 188; Mismatches 492; Indels 290; Gaps 39;

QY 56 PRIIEPTDVLWKNKPNATLNCVKBGKPBPTIEWFKDGEPVST--NKKSHRVQFDGAL 113
DB 68 PRIIEPTDVLWKNKPNATLNCVKBGKPBPTIEWFKDGEPVST--NKKSHRVQFDGAL 113
QY 114 PFYRTMKQKQEG--DGEWCVARNVGVQSHVASQIAVLDRFVPEKDTVAKGETA 172
DB 128 PFYRTMKQKQEG--DGEWCVARNVGVQSHVASQIAVLDRFVPEKDTVAKGETA 172
QY 173 LLECGPKGPIEPTLIWIKDGVPLDLAKMGASGRVIRVDGNLISNVEIDEGNYK 232
DB 188 LLECGPKGPIEPTLIWIKDGVPLDLAKMGASGRVIRVDGNLISNVEIDEGNYK 232
QY 233 CIAQLNLTRESSTAKLVQVQKPYFMKEPKDQVLMVGQTAIFHCVSGDPPPKVLAKKEE 292
DB 240 CYNMGVGESEVAELTVLERPSVLRPSNATVTVDSAEAKCEARGDPTVIRWDD 299
QY 293 GNTPVSARILHDEKSLNISNPTDEGTYVCAHNNVQVLSARASLIVAHNPKYKRPS 352
DB 300 GNTPVSARILHDEKSLNISNPTDEGTYVCAHNNVQVLSARASLIVAHNPKYKRPS 352
QY 353 NKKVGLGVQVLPCHASGNPPSVFVTKRGVSTLMF--PNSSGRQTVAAQDTQITDV 409
DB 359 NKKVGLGVQVLPCHASGNPPSVFVTKRGVSTLMF--PNSSGRQTVAAQDTQITDV 409
QY 410 RQEDSEGYTSCAFSVSDSTVRVQLQVSSV-DEPPPIIQIGPANQTLQKSVATLPCRA 468
DB 419 RQEDSEGYTSCAFSVSDSTVRVQLQVSSV-DEPPPIIQIGPANQTLQKSVATLPCRA 468
QY 469 TGNPSRIKHGHGAVQA-GNRYSIQGSLLVDQLSDSGTITCTASGERGETSWAA 527
DB 479 TGNPSRIKHGHGAVQA-GNRYSIQGSLLVDQLSDSGTITCTASGERGETSWAA 527
QY 528 FLTVKEKP-STSLHRAADPSTYPAPPQPKVLNRSKTSISLWAKSKQKPGAVGLIQT 586
DB 539 FLTVKEKP-STSLHRAADPSTYPAPPQPKVLNRSKTSISLWAKSKQKPGAVGLIQT 586
QY 587 VEYFSPDLQGTIVAAHVGDTQVIGSLTPTGTVFLVAENQICGVSPGLSNVIRTI 646
DB 596 VEYFSPDLQGTIVAAHVGDTQVIGSLTPTGTVFLVAENQICGVSPGLSNVIRTI 646

QY 647 EADFDAAASAMLSAARTLLTKRSELVIDASATNASAVRLENLWHSVDEKIVYGLRIHYK 706
DB 656 dv-lptsgvdkhqqvrelgnvhlhmtplssssievhwt--vdqsgyqygykilyr 712
QY 707 DASVPSAQZBS-----ITVMDASAEFVGNLAKKVKYEFPLFPFETIEGQPSNKTA 760
DB 713 -----paganhgesdvlfverrtpaknsvipdlrkvnyelkarpfnefgadseikfa 768
QY 761 ILYEYDPSAPDNTQIGMT--NQTAGVWMTPPPSQHNCLNYKIKTIEVSAGNTMKVAN 818
DB 769 kileaspsappgvvtvskndgngatlsvsgpppdtqgmqgqv-wclngetryhin 827
QY 819 MILNATTTSLNLLNLTGAVTISVRLNSFTACRQGPYSKPSISLMDPTPHVHPRAHPSGT 878
DB 828 ktvgdtfsvfpilvpglrysvveaasgagsgvsksepfqld-----ah----- 874
QY 879 HDGRHSGQDLTTHNNGN-IPPGDINPTTHKTTDYLSP-----WLMVLVCIWLL 927
DB 875 -----gnpsvped-qvslaglsdvkqpfaglagiaacvillwfsiwl 918
QY 928 VLIVISAISNIVYFKKKH-----MKTELHLVSDNEITALNI 966
DB 919 -----yhrkRngltstygarkvpsftfptvtvrggeavsggrpdlm 967
QY 967 NKSSEL-WI-----DHRGWRTADTKDGLSKSLKSWNSQ--SNWNS----- 1010
DB 968 sepaagpvladtvgptgnhndcsiscatgngnsdnltyrpadciaynqndlqk 1027
QY 1011 -----DGGDTVAEVDNLTNLTFCNCRKSPD-----NPTPYATTMIIGTSSECT 1054
DB 1028 tnlmpestvtygvdlnkinemktfnspnlkdgrfvnsgpqtpyattqlqslsnm 1087
QY 1055 THTTISADK-----DSGTH 1065
DB 1088 nngsgdsgknhkplgqkqevapgyvaiveqnlkdyrandvtpetipynsgqnt 1147
QY 1070 SPISDAFAGQVPAVVPVVKSNLYQVPEP-----INNSFLPPPPPPSPSTGYAQSPE 1125
DB 1148 gysnnsgdsgtsgsqghkgartpvpkqgmawdlpppppppphps----- 1197
QY 1126 SRRKSKSAGSGISTNSILNASIHSSSGGFSWAGVSVQYAVACPPENVT----- 1176
DB 1198 -----nseynlsvdes-----ydgempcpvparmylqdelee 1233
QY 1177 -----SNPLSAVAGTQNRVQITPTQNH--PPQLPATYATTGGGAVPNH 1220
DB 1234 edergpttpvrgaasp-aavsyshgstatltpsqeqlgmldcpeet-----h 1284
QY 1221 LPFATORHAASETOAGLNAARCAQSRACNSDALATSPQKPPPPVPPBQWTVPHNS 1280
DB 1285 nqhgqdr-----qpsvpppp-prp-----lsphty 1312
QY 1281 HPMHPTSSN 1289
DB 1313 yisgplvd 1321

RESULT 6
Y13565
ID Y13565 standard; Protein; 1297 AA.
XX
AC Y13565;
XX
DT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX
PN W09925833-A1.

XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
XX
DR N-PSDB: X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.

The invention relates to a method for modulating the amount of Comm (commisuraless) polypeptide in contact with a cell expressing active Robo (roundabout) on its surface. The method comprises modulating the effective amount of Comm polypeptide in contact with the cell, where the amount of expressed active Robo is specifically modulated inversely with the modulation of the effective amount of Comm in contact with the cell. The method is used to modulate the amount of active Robo expressed on a cell. The method can be used to screen for agents that modulate Robo:Comm interactions. This is particularly useful for modulating nerve cell function.
XX
SQ Sequence 1297 AA;

Query Match 21.4%; Score 1588; DB 20; Length 1297;
Best Local Similarity 31.0%; Pred. No. 3.8e-83;
Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

QY 55 SPALIEHPDLDLVKNEPATLKNKVGKPEPTLEWFGDEPVSNEKK--SHRWQKDG 111
DB 29 SPVLEHEDLVKNEPATLKNKVGKPEPTLEWFGDEPVSNEKK--SHRWQKDG 86
QY 112 ALFFPTMGK--KEQDGEWCAVNRVGVASRAHSIAVLARDFRPEKDRVAG 169
DB 87 SIFILKVASGKNGKSDAGAYCVASNEHGVKSNEGLKILMREDFRVPRTVQLG 146
QY 170 ETALLCGKPGPIPEPLIWKDGLVDLKLKAGSSSRVIVDGNLINSVPEIDG 229
DB 147 EMALCECPGPIPEPVVSRKDD----KELIQDPRTYIHSNGNLIIDPDVRSAG 200
QY 230 NYKICIAQLWQVRESSAKLIVQVPRYKPKQDQVLMGQTATPHVSGDPPVLMK 289
DB 201 TYGVANNVGVSRVSPARISFEKPKFEQPKMTVDVGAVALICDRTVQPPGQITW 260
QY 290 KEBNGPIVSRARILHDEKLEISNITPDEBTVCAHNNVQGISARASILVHAPPNFK 349
DB 261 KNEPMPTRAYIAKDNRIERIVGSGDEGEVYCARNPAGTLEASHLVRAGPSGT 320
QY 350 RPSNKKVGLNGVQVLPQMASGNPPSPVFTKGVSTLMFNP--SSHGRTVAADQLTIT 407
DB 321 KPADGVPAGGTATFECTIVGSPSPAYFSWKGQGLIFPSYSDGRTKVPSTGTITLE 380
QY 408 DVQKDEBGTVCYSASFVDSSTVRFVQLVS----- 437
DB 381 EVRGDEGAYVCAQMSAGSSLSKAALATFETKGRVQKKEKKGKQKQKVSILKYL 440
QY 438 ----SVDERPPPIIQIGPANTLPKGSVATLPKATGNPSRIKWFHGDHAVQ--AGNR 491
DB 441 SAVGTNPKAKPTIEHGHQNTLVGSSALPCQASGKPTGISWLDIGLIDITDRI 500
QY 492 SIQSGSLVNDLQSDSTYTCASGERGTSNAALILVTEKPGTS--LEHRAADSTPYA 550
DB 501 SQHSTGSLHIAIDKPDGTGYTCLAKNEDGTSVALSVEDTSNAQVIRMPDQNFPS 560
QY 551 PPTPKVLVNSRTSILSWAKSGEKRGAVGPIIGYTVFSPDQITGMVAAHVGQDQV 610

DB 561 SPTQPLVNVDTETEVLH--NAPSTAGPITGYLIGYSDPLGQWFLNPDYVASTAY 618
QY 611 TISGLTPGTSVFLVRAENTQGISVPSGLSNVKTIEADPDAASAN----DISAARTLL 666
DB 619 RIKGLKPSHYMFVIRENEKIGTSPSSALVTTSKPAQVALSDKNMDMAIEKRIT 678
QY 667 GKS-VELIDSAKNAKSAVLEWMLHVSADKTYEGLRIHYK--DASVPSAQYHSITWDS 724
DB 679 SEGLIKLEVKTIKSTAVRIFWKRKL--EELIDGYIKWGPSTNDQNG--VQVTS 734
QY 725 AESFVGNLKKYTYEFPLTF---PFIIEGQPSNKTALTYEDVPAPPDQIWMQNG 781
DB 735 TENYVSLMPTGYEFFVYHSGVHSIGAPNSMDVLTAEAPPDPEVIRIMNL 794
QY 782 TAGVWNTPPPSQHHNGMLYGIKIEVSAGNTMKVLNNTLNAATTSVLLNLTGAVIS 841
DB 795 TLRISKAPKADGINGILKGFQI-VIYGAPANNRNTIATRAASVTLFHLVTGYTKI 853
QY 842 RLNSFTKAGDGYKPKISLMD-PIHHVHPRAHPSDQDRHGDQLTTH--NGNIPP 898
DB 854 RVARSNGGYVSGTSEVLMQDLEKHLA-----AGGENSFYIYLINKSHRP- 903
QY 899 GDNPTTHKKITVTLSPWMLVLCIVLLVISAISVYFKKHKMKEGLHGSVSD 958
DB 904 -----VIVALLIFVILAIYCYWNSRNSDQKDRSIFKIND 942
QY 959 NEITALLNINSKESLW----IDHHGWTATDTKDSGLSEKLLSHVNSQSNVNSD- 1011
DB 943 GSVHMAN-----LWDVAQNGQPNMYTAGTMMNRGQALYIPLAQDFFNCDDY 998
QY 1012 GGT-----DYAEV--DTRNLTFVCKRSPDNPTATWTLIGTSSSECTKT 1058
DB 999 AGTHMPSEHGHYAGLTGGPNANSTFY-NGYIDGSPYATTLV----- 1045
QY 1059 SISADKDSGTSPSDFAGQVPAVFNKSNLYKQIPEVINWSEFLPPPEPPPSSTG 1118
DB 1046 -----LNGQQA--WLDNMLRAPANTN-----PVPE--PPRYED 1079
QY 1119 YAGSGPSSRSGSSKAGSISTQSLMASHSSSGSGFAGVSPQVATPCAPVENYSN 1178
DB 1080 HTEG--RRSRASRAGDQ-----TLNGILHRTSGQRS-----DSPPTDVS 1122
QY 1179 PLSAVAGGTQNTYQITPTPNQHPOLAPY-FATTPGGAQVPR-NHL-PPATQRAASETA 1235
DB 1123 VQLHSAGDTGSKERTGERTPGNKLMDFIPPPPNPPGGGHVYDATTGQ----- 1175
QY 1236 GLMAACQASQACNSDALATPSPMQPPVPEVQGVYVPHNSPHMPTSSNHIQTC 1295
DB 1176 -LARGSTPRETDYS-----VSDQAFADVNA-----RPTSRNRL- 1211
QY 1296 SSECSDHSRSGSKKQLQLEHSGKASQKRGHRRAP 1334
DB 1212 GGRPLKGRDDSGSSLMDDDGSGSEADGENSGDVP 1250

RESULT 7
Y08403
ID Y08403 standard; Protein; 1297 AA.
XX
AC Y08403;
XX
DT 24-JUL-1999 (first entry)
XX
DE C. elegans ROBO protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Caenorhabditis elegans.
XX
PN W09920764-AL.
XX

```

XX RESULT      8
XX Y08404
XX ID   Y08404 standard; Protein; 1649 AA.
XX XX
XX AC   Y08404;
XX XX
XX DT   24-JUL-1999 (first entry)
XX XX
XX DE   Human ROBO1 protein.
XX XX
XX KW   ROBO1; ROBO2: roundabout; nerve guidance; human; murine; cell function;
XX KW   cell morphology; screening assay.
XX XX
XX OS   Homo sapiens.
XX XX
XX PN   W09920764-AL.
XX XX
XX PD   29-APR-1999.
XX XX
XX PF   20-OCT-1998: 58WO-US21264.

```

XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-112615/26.
DR N-PSDB; X08404.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 65-71; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1649 AA;

Query Match 21.34; Score 1584; DB 20; Length 1649;
Best Local Similarity 30.14; Pseq. No. 9.1e-83;
Matches 418; Conservative 188; Mismatches 493; Indels 288; Gaps 39;

Qy 56 PRILEPTDVLWKKENPATLCKVEGKEPTIEWFKDGEVPT-VEKSKSHVQFDGAL 113
Db 68 PRIVEGSLIVKSGEPATLCKEAGRPTEVYKSGSERVEDKDPGRHMLPEGSL 127
Qy 114 PFYRMQKKEK-DGSEWCWAKNRVQVGRASLQJAVLADDEVPKDTVARGETA 172
Db 128 flfvlrvksgprdegvyvavnylgeavshmslevallddrgpsdvavveqpa 187
Qy 173 LLEGCPGPIPEPTLWIKDGVPLDLKMSFGASGNRVVDGGLLSNVEPDEGNK 232
Db 188 vmeqcpgrpbeptiswkdgspld-----kderiti-rygkintytrksdekyv 239
Qy 233 CIAQLNWTGRESTAKLIVQKVPFMKEPKDQVWLIGQTATFHCVGDDPPVKLKKKE 292
Db 240 cyqtmwgereseaelvlerpsfvkrpsnlavtvddseefcoearpdpvtrwkd 299
Qy 293 GNTPYSRARILHDEKSLSEISNPTDEGTYCEAHNRWGOISARASLWAPNFTKRPS 352
Db 300 gelpkyryel-rdhtlklrvktagmgytoavemvykkaesatlvepghfvkrp 358
Qy 353 NKKVGLGVVQLPCMASGNPPSVFWTKRGVSTLMP-----PNSHGRVNAQDTLQITD 409
Db 359 dqvalrvrtvtgqeatgnppaifvrregslifysqppgsrsfsvagtqdlctav 418
Qy 410 RQBDEGTYVCSAFVSDSIVRVFLQVSSV-DERPFPPIIQGIPANQTLPKSGVATLPCRA 468
Db 419 qrsdvgyllqclnvagsiltkeylevtdvldppvrlqgppvqvtavdgfvfscva 478
Qy 469 TGNPSRIKWFHDGHAVQA-GNRTSIIQGSLLRVLDQLSDGTTCTASGERGETSWAA 527
Db 479 tsgpvtlwlrvkdvlvstgdsklenglyqlryaklgydtrytclastpsgeatwaa 538
Qy 528 TLTYKFGP-STSLHRAADPSTYPPAGTPKVLNRSISLRLWAKSGEKPAGVPIIGIT 586
Db 539 ylevgefvyvpprrdpnlpisapskpevtvsvrntvlw---qnlngatpsty 595
Qy 587 VEYFSPLDQGVNVAHRVGDQVQISGLTPTGSYVFLVRAENTQGISVPSGLNWKTI 646
Db 596 leafshaagsgvqtveenvktetsaiklpnalyflvraanaygidspsqlsdvkvtk 655
Qy 647 EADDTNANANDLSARTLLTGKSVELIDASAINASVRLWMLVSADEKTVBELRIHK 706
Db 656 dv-lptsgvdkqvqrelgnavhlhptvlsssslewhv---vdqsgyilgykilyr 712

Qy 707 DASVPASQYHS-----ITVMDASAESEFVGNLKKYTKIEFFLTPFPETIEGQSNKTA 760
Db 713 -----peganhgesdwlfevtrtpaknsvipdlrkgyvnelkarpffnefgadseikfa 768
Qy 761 LTYEDVSPAPPANOIQLMY-NQTAGWVRWTPPPSHHNNLNGYKLEVSAGNTMKVLAN 818
Db 769 ktleaspappggvtskndgntailvsvqpppedtqngvqeykv-wclgnetryhln 827
Qy 819 MTLNATTSVLLANLTLTGAVTSVRLNSPTKACGPGYKSPKISLPMOHPHVRPRAFSPT 878
Db 828 ktvdgetsvvplflvpglrysvaastgagsgksepqfild-----ah---- 874
Qy 879 HDERHGEQQLTNNYNN-IPPGDINPTTKRTTKYLKSGP-----WMLWLVCIVLL 927
Db 875 -----gnpvspe-d-qvslaqlsdvkvqafilagiacvillmwfsilw 918
Qy 928 VLIVISAISIMVYFKRKHQ-----MTKELGHLISVSDNEITALNI 966
Db 919 -----yhrhkrknegtstyaqirvpsfiftptvyrqgeavssgprgpllnl 967
Qy 967 NKESL-WI-----DHEHGWRATADKDKSDLSKSLISVNSSQ-----SNTNYS 1018
Db 968 sepaagwladvptgnnnhdscisctagsgnssnltyrpadclnynmqldnqk 1027
Qy 1011 -----DGGTDAVEYDTRMLTTFYVCKRSPD-----NPTPYATTMLIGTS 1049
Db 1028 tnlmlpestvygdvlsknknemktfnsplkdgrfvnpgsgptpyattlqslslnmnn 1087
Qy 1050 -----SSBCTKTTISADKSDGTSHP 1071
Db 1088 gsgdsgekhkwplggqkqevapqynlveqnlkldyrandvtpflynpgsqdqtggs 1147
Qy 1072 YSDAPQVQAPVPPVKSNTLYQVPEP-----INWSELPPLPPHPPSTYIGACSPSS 1127
Db 1148 ynsdrgsgtsgsgqghkgartpkvpgagmmvadlppppahpshs----- 1195
Qy 1128 NKKSGAGSGISNTQSLNASHSSSGSGPAGVSPQVAVACPENNY----- 1176
Db 1196 -----nseeylsvdes-----ydeqpcpvypparnylqqdeleed 1233
Qy 1177 -----SNPLSAVAGGTQNTQYITPNNH-PPOLPATFATTPGGGAVPNWLP 1222
Db 1234 erqtpvprvgassp-aavysghgstaltitpsgeelqnlqdcpeatg-----hmq 1284
Qy 1223 FATORHAASSTQGLNARCAQSRACNSDALATPSMPQPPVPPVPGWGTQVPHVNSHP 1282
Db 1285 hqpdrrr-----qpvsppp-prp-----isphttyyl 1312
Qy 1283 MHTPSN 1289
Db 1313 sgplved 1319

RESULT 9
W83927
ID W83927 standard; Protein; 753 AA.
XX
AC W83927;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human T85 protein.
XX
KW T85; FHMG-5D4; PMWV-SD4; human; neurological disorder; therapy;
KW diagnosis.
XX
OS Homo sapiens.
XX
PH Key location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Protein 21..753


```

RESULT 11
W42086
ID W42086 standard; Protein; 1910 AA.
XX
AC W42086;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human Down syndrome-cell adhesion molecule DS-CAM1.
XX
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis: assay: human.

```

Query Match 8.9%; Score 661; DB 19; Length 1910;
Best Local Similarity 24.2%; Pred. No. 1.4e-29;
Matches 305; Conservative 175; Mismatches 473; Indels 308; Gaps 60;

Qy	55	SPRIIEHLDVLVKKNEPATLVCBKGPPEPTIEFWKQGPVSTNEKSSHQV-----FKDG	111
Db	406	tpklissfsekvspaeplslnmncvqtpptitwlddopli-----kggshrismitge	463
Qy	112	ALFFETITQGGKQDGGYTCVANKNRQGVASRHASLIQVLADRFVPE-KDTRVAKGE	170
Db	464	nvvsylnlssgqvrdvggyrcrtannsg--vlygqrliv-----rpspsrlnmkntatag	519
Qy	171	TALLCEGPKPKIPEPTLIWKDG--VPLDLMKAMSGASRRVIRVGGNLLISVW-PID	227
Db	520	dttyhc-tvlygpyysiklyknsollpf-----hrgvafennqtklsvdqvkd	569
Qy	228	EGNYEK--IAQLNLWGTRSSYAKLIVQKVPF--MKPEQDQVNLVQGTATFHC-SVGGD	282
Db	570	egeytcnvlvpqlstlsgvsh--vtvkvpfpflgpfefpfrfsi-----qgrvfipcvvsgsl	624
Qv	283	PPKVLAKKEENIPVPSRAILHD-----PKSLIESNTPTDGTWCYCAKNVAGUISAASL	339

Db 625 pttitwqgdrpipsylvtdnidftssicnslmnhgnyctiarneaaevhsqgl 684
 Qy 340 IVHAPPNFKRPSNKKVGLNG-VVQLPCMASGNPPPSVEWT-KEGVSTLMF-PNSSHGRQ 396
 Db 685 lvrvpdkfvvqprgdg-glygkavilnceagaypvtivwkskagvpgqfplnqgr 743
 Qy 397 TVAADGTLQITDVQEDEGTYVCSAFVSDDSTVR-VFLQVSSYDERPPPIIQIGPANQT 455
 Db 744 qvlngslilkhvveedsgyllckvsnvqgadskemyltv-----kipamitsypttll 798
 Qy 456 LFKGSVATLPCRAATGNPSRIKWFEDGHAQVQAG-NRYSIIQG-----SSLRWDLQLLS 507
 Db 799 atqgqkkmstahgkpllvrvkedrlloepmarylvstkwgeevietlqlptvre 858
 Qy 508 DSGTITCTACGERGERTSWAATLVEKPGSTSLRAADPSTYPAPGPTKVLNVSKTSL 567
 Db 859 dsqfshcainsygedrglqltqvpe-----pdpeel-ekdkvartitl 903
 568 SNKASQEKPGAVGPTIGTYVEFSPDLQTQWIVAARHVD-----TQVTSIGLTPTSTTV 622
 Db 904 rtwmgfd--gnspitgydie--cknkadsvdsagrtkdvspqnsatlidhpsstys 957
 Qy 623 FLVRAVNTQGISVPSGLSNVIXITLQADFDASANDLSAARTLLGKSVLEIDASAINASA 682
 Db 958 imyknriksgsep--snel-titad-easp-----dgpqqe-vhlepissqs 1000
 Qy 683 VLEENLHVSADEKT-----VEGLARHYKQASVPSAQYHSITVMDASABS--FVNVNLRK 735
 Db 1001 lrvtw-----kpkkhlgqngirgylgyreystggnfgnliavdsgdeevytdlnk 1056
 Qy 736 TYKEFFFLPFPETIEGQPSNKTALTVEDVPSAPPDNIQIGMTAGWVWTPPPSQH 795
 Db 1057 fctglvlvqacmrastgssgeftittdelpaypenvgalatssesiswstlskea 1116
 Qy 796 HNGNLGKHYKIEVSAGNWKMLANMTLNAATT--SVLLNLTGAVISVRLNSPKAGDGP 853
 Db 1117 lnglgrfv-lywanlmdgelgetknittqpsleldiglektynsyqlaftragdsv 1175
 Qy 854 YKSPISLMDPTTHVHPRAAPSCGTHDGHGQDLTYHNGNIPPGDINPTTHKTKTDTL 913
 Db 1176 rseql--ftrtkedvppp--pag----- 1194
 Qy 914 SGWMLVLCVILLVLVISAISMVYFK-----RKQMTKELGSLVSDNE-- 960
 Db 1195 -----vkaaaasasmfvewlppklnglirkytvfcsghpytvlsefas 1240
 961 -----ITALINNSKESLWIDHHRGWRATDQKSGLESKLLSHVNSQSYNNNSDGG 1013
 Db 1241 pdcfsyrlpolsrurgysvw--vavtsagry-----nsseil 1276
 Qy 1014 T--DYAEVDTKRLT-----TFYNCRKSPDNPTPIATTM--LIIGTSSETCTKT 1057
 Db 1277 tveplakapariltfsgtvtppmklvlpv-kavgdpepavkwmkdsngtpelvtidgr 1335
 Qy 1058 TISASD-----KDSGTHSPYSDAFAGQVPAVVPVSNLYQFVPEPINSE----- 1102
 Db 1336 rsifsgsfliirtvkaedsgyys-----clann-----nwsdelil 1372
 Qy 1103 --FLPPPPHP-----PPSSTYGYAQGSSPSSKSSKAGSGISTQSLNINSHSS 1152
 Db 1373 nlqvqvppdqplrtvaktssitlslwpd-----nqgsseirgylqysedns 1421
 Qy 1153 SSGGFSANVSGVQAVACPPENYVSNPLSAVAGGTQNYQITPTNQRPQLPAYATTGP 1212
 Db 1422 eq-----wgsfp-----ispersyzt--lenlkogtykftitagn-----gwpp 1459
 Qy 1213 G 1213
 Db 1460 G 1460

R13144
 ID R13144 standard; Protein; 1728 AA.
 XX
 AC R13144;
 XX
 DT 04-OCT-1991 (first entry)
 XX
 DE Deleted in Colorectal Carcinomas.
 XX
 KW DCC gene; cancer; diagnosis; antibodies; tumorigenesis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Protein 202..1648
 FT /label= DCC
 FT Peptide 202..227
 FT /label= sig_peptide
 FT Protein 228..1648
 FT /label= mat_protein
 XX
 PN W09109964-A.
 XX
 PD 11-JUL-1991.
 XX
 PF 19-DEC-1990; 90WO-US07314.
 XX
 PR 04-JAN-1990; 90US-0460981.
 XX
 PA (UOJO) JOHNS HOPKINS UNIV.
 XI
 PI Vogelstein B; :
 XX
 DR WPI; 1991-222913/30.
 DR N-PSDB; Q12752.
 XX
 PT Human DCC gene, deleted in colorectal carcinoma - and diagnosis
 PT or prognosis of neoplasms by detecting loss of gene function or
 PT expression prods., or mutation(s)
 XX
 PS Claim 44; Page 31; 51pp; English.
 XX
 CC Cells transformed with the wild-type DCC gene can be used as model
 CC systems to study cancer remission and drug therapy. DCC polypeptide
 CC expression prod. may be used to reverse the neoplastic state.
 CC X1615 represents an amino acid illegality in the specification, all
 CC other Xs are encoded by stop codons.
 CC See also Q12752*55.
 XX
 SQ Sequence 1728 AA;
 :
 Query Match 8.6%; Score 639; DB 12; Length 1728;
 Best Local Similarity 23.1%; Pred. No. 2.2e-28;
 Matches 361; Conservative 170; Mismatches 565; Indels 464; Gaps 70;
 :
 Qy 57 RIIEHPDLVKKNEPATLNCKVKG-KPEPTIEWFKDG--EPVSTNEKKSHRVQKDGAL 113
 Db 242 rlfesepdavnrgnvnildcsaesdrpvykvkwdghlhalgnderkq-----qlsngsl 298
 Qy 114 FFYTRTQKKKED-DGGEYKCVAK-NRVQAVSHASGLQIA-VLRDRFVRPEKDTVRKAGE 170
 Db 299 lliplllhrhkhkdeglyqceasldsgslisrtakvavglpr--flagstevtafmgd 356
 Qy 171 TALLEGGPGKPIEPTLTIKDGVPDLDMKMSGASSRRVRIVDQGNLLISNVEPIDEGN 230
 Db 357 tvllkc-evlsepmttlhwkqgqdltpip-----gdsrrvlpssglqisrlqgldgl 410
 Qy 231 YKICINQLNIGTRESSYAKLIVQVKE-----YENKPEKQVMLIGQATPHCSVGGDPPP 284
 Db 411 yrcsarnpasstrgneavrilsdgplhrqlfqlgrpsnvaiegkdavleccvsyggpp 470
 Qy 285 KVLWKEDEGNIPV-SRARILHDEKSLISNITPTDGGTYVEAHNNQGISARASLIYHA 343

471 sftwlrgeavlglskysllggnllnsntddsgmtyctvtykneilsaeelvlv 530
 344 PPNFTKRPNKNGVLGVQLPCWASGNPPSPVFTWKGVSTLMPNNSHGRQTVAAEDT 403
 531 ppwflnhpsnlyayessndlefectvsgkpytpvnmknq---dvrlpsdyf---qlvggsn 585
 404 LQITDVRQEDGTYTCSAFSVDSSTVFRVLQVSSVDERPPPIIGFANMTPKNGSVAT 463
 586 Lrlglvksdgyfyqvaenaagtaeqllv-----pkpalpsssvlsaprdvvpv 639
 464 L-----PCRATGNFSPRIKWFHDGHVAGQNRYSIIQ-----GS-SLRWDDQL 506
 640 lvsrfrvlrsvpaaekgn---lgtfvffaregdrreralttgpgqltlywnlkp 695
 507 SDGTTTCTASGERGETSWAATLTVEKPGSTS---LHRAADPS-TTPAPGTPKVLNVSRT 563
 696 eamtyfrvaynewg-----pgessqplkatqelqvgpvenlqvastsept 743
 564 SISLWMAKSGKPG-AVGPILIGTVYFVPSDLQTG----- 597
 744 sliltv---eppayangpvqgyrl---fctevstgkeqnevdlsgyleklkftteysl 797
 598 NIVAAHEWGD-----TQVITS----- 613
 798 rflayrnpygysdtditvvlsvdpsappgvslvwnrsikswlppesqtqngfit 857
 614 -----GLTPTQTSVFLVRAENTQGISVPSGLSNVKT 645
 858 gykirhkrtrrgemetlepnwllyflgksgysgysvaantvngtppsnwyta-et 916
 646 IEADPDAASANDLSAA-----RTLLTGKSV-----ELID 674
 917 pendlgesqvgdpslhrpqtncimsvtpplnpavrryglygvgspayetrnd 976
 675 ASAINASVRLNMLHV-----SADEKTV----- 698
 977 skqryslysterlessshyvislkafnaggevgpylesatrsttdptdvdypllddft 1036
 699 -----BGLRIHYKDSVP-----SAQY 715
 1037 vpydltspmlppvgqavalthdavrsvadnspqkntseavlytrvrtfsasaky 1096
 716 HSITVMDASAESFVGNLAKYTKYEFLTPFFETIBEQSPSKALTITEDVP-SAPPDNI 774
 1097 ks---edtlslsyatgktpnteyfsvmtnkrrestvmsahattyeaptsapkdf 1153
 775 QIQHTQG-TAGWVRWTPPPSQBHNCNLGYKIEVSAGNTMK-----VLAMTLNATTSVL 829
 1154 vitregkpravivewqpp--leangkitayllfityldkplddvmetlsgrlthgim 1211
 830 LNNLTGAVISVRLNSFTKAGDGYSKISPLMDPHRY-HPR-ARHSPGS-DGRHGG 886
 1212 dldldt---myrflqarnsgvypldpi-lfr---tlkvhpdkmandgthgdgypwv 1266
 887 D-L-ITNENGNIPP-GDINPTHKTKTDYLSGFWLWLV-----CIVLWLVSAAL- 935
 1267 dtdltdrlsteppigqmh-pghsvtpqksnllvityvtvtyvlvrlvavicttr 1325
 936 SMYVFKR---HMTKELGHLVSVDNEITALMINSKSLMDHHRGWTADTDKSG- 990
 1326 ssaqqqrklrthasgkrkg-----sqkl-----rppdlwi--heememknlekpegt 1373
 991 -----LSEKLSLHVNSQSQMYNNSDGTGYAEVTDTRNLITFYCNKRSQDMPTFYATT 1043
 1374 pagrdapligsqdltvphsqsetqlgskstshgqdtce----- 1413
 1044 MlTGTSSETCKTKTTSISADK-----DSGTHSPSDFAGQVAVPVVXSNLQY 1093
 1414 -----agssmetlerslaarreprkrlnipmdagsnpp---avrslyptvlesaqy 1462
 1094 PVEPIMNSFELAPP-EHPPPSSTY-----GYAOGSPSSRSKSSKSGSGIST 1140

Db 1463 P-----gllpsptogyphqftrlpvfpvlsvdrfgag-----rsqsvseqpt 1508
 QY 1141 NQSL-----WASHSSSSSGGFSAGVSPQYAVAC-----PPENVISNLSAVAGGTQNRQY 1193
 Db 1509 qqqpmlpsqpehessee-----asprstactvcrphtlrsfanllp----- 1552
 QY 1194 TPTNORPPLPFAFATGTGGGAVPVPHNLPFATQTHAASQYAGLNNARCAQSRACNSCA 1253
 Db 1553 pmsalekvvypplsgpgptlpkthvk-----taslglagkars----- 1593
 QY 1254 LATPSMQPFPVVPVPEGWYQPVPHNSHMHFTSSNNHIOCSSESDHSSSSQSHKROL 1313
 Db 1594 -----pllpvsvpta--perseesh--kptxdsanrye-qdlseqmasleglmkl 1640

RESULT 13
 R68553
 ID R68553 standard; Protein; 1447 AA.
 XX
 AC R68553;
 XX
 DT 05-JUL-1995 (first entry)
 XX
 DE Deleted in colorectal carcinoma (DCC).
 XX
 KW Tumour suppressor; deleted in colorectal carcinoma; antibody;
 cancer diagnosis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..1063
 FT /note= "DCC epitope"
 FT Misc-difference 3369..4341
 FT /note= "DCC epitope"
 FT Misc-difference 26..1126
 FT /note= "DCC epitope on extracellular domain"
 FT Misc-difference 1123..1447
 FT /note= "DCC epitope on intracellular domain"
 XX
 PN W09428161-A.
 XX
 PD 08-DEC-1994.
 XX
 PF 18-MAY-1994; 94WO-US05277.
 XX
 PR 26-MAY-1993; 93US-0068950.
 XX
 PA (UWJO) UNIV JOHNS HOPKINS.
 XX
 PI Bruskina A, Jarosz DB, Johnson K, Kinsler KW, Vogelstein B;
 PI Zaretsky JR;
 XX
 DR WPI: 1995-022830/03.
 DR P-PSDB: Q80196.
 XX
 PT Antibodies specific for tumour suppressor gene product, DCC -
 PT useful for detecting expression of DCC gene, for cancer diagnosis.
 XX
 PS Claim 4; Page 24-28; 3pp; English.
 XX
 CC The protein represents the DCC tumour suppressor, and epitopes are
 identified which are used in the generation of polyclonal or
 CC or monoclonal antibodies against DCC. The antibodies can detect
 CC DCC protein in biological samples (including tumour tissue,
 CC peripheral blood mononuclear cells or a tumour biopsy lysate)
 CC despite low levels of DCC expression, and are therefore useful in
 CC cancer, especially colorectal carcinoma, diagnosis.
 XX
 SQ Sequence 1447 AA;

Query Match 6.6%; Score 637.5; DB 16; Length 1447;

Qy 57 RIIEPTDVLVKNPEATLCKVEG--KEPTIEFWKDG--EPVSTNEKSHRVQKDAL 113
 Db 41 flfsepsdvtmrgrnvlcdsaesdrvgpvkvvkqqlhalglenderkq---qlngsl 97
 Qy 114 FFPTIMQKKEQ--DGEWCVAK--NRVQVARSVHSAQIA--VLRDFRVKPKTRVAKGE 170
 Db 98 lllqllrhkhhkdeglyqceaslgydsgslisrtakavagplr---flsqetvafmgd 155
 Qy 171 TALLDCEGPKGPIPEPTLWIKDGVLDLKAWSFGASSRVIVDGNLLINVEYIDGN 230
 Db 156 tvllke-evlgepmthlwqngqdltpip-----gdsrvvlpesalqlslrpgdigi 209
 Qy 231 TKCIATMNLVGRSTSTAKLIVQVKP-----YPMKPKDQVWLYGQATFHCVSQGDPPP 284
 Db 210 yrcrsanpassrtqneevrlsdgplhrqllyflqrpsvvaiegkdvleccvsgypyp 269
 Qy 285 KVLAKHKEGNIPI--SRARILHDEKSLISNITPPDGETVYCEAHNVQISARASLIVHA 343
 Db 270 sftvlrgevlqlrskysilgssnllsvtdsdsgmytctvtykamsasaelv 329
 Qy 344 PPMFTKRPNSKVKVGLGVQLPCASGNPPSPVSTLMFKNSSRGRTVAAGDT 403
 Db 330 pwwflmhsn-----lyaysem----- 346
 Qy 404 LQITDVRQEDGETVCSAFSVSDSTVRVFLQVSSYDERPPPIQIGPANQLTKSGVAT 463
 Db 347 ----dlfe----- 351
 Qy 464 LPRCARGNSPIKRWFDGHAVQAGNRISLIQSSILVDLQSDSTYTCTASGERGET 523
 Db 352 -ctvsgkpvptvmwmgndvipsdyflqvgssnrlilgyvksdegfyqcvaeag 409
 Qy 524 SWAATLVKPKGSTSLRAADSTYPAPPKTVKLVNRSSTLSLWAKSQEKPGANGPII 583
 Db 410 qtsaqilvlpkpaips-----ssvlpsaprdvrvpvlsvsrfvrlwppae---aknglq 460
 Qy 684 GITVETVSPD-----LQTGWIVAAHRVGDQTYTGLTPGSTVTVLVAENVTQGISVP 636
 Db 461 tftv-fisregndreralnt-----tpqslqltvgnlkpeamtyfrvaynewg--p 510
 Qy 637 SGLSNVITIBADFDASANDLSAARTLLTKGSVELIDASAINAVLEWMLVSADEX 696
 Db 511 gessgpiq-----vatopelqvgpvenlqavstpsitlilevppayangp 557
 Qy 697 TVBGLRIHYKDAVSPSAQYBSITVMDASAEFVGNLKKYTKYEFFTLTPFETIDQSPN 756
 Db 558 -vgyrlfcterstgkegn-----levdglaykleqkkfeyslrfleyncrypgsvtd 611
 Qy 757 SKTALTEVDVSPAPDNIQIMYNQTAGVWYTPPPSQHNNKILGTIKI-----EVSAGNM 813
 Db 612 dtivtclsdvpsapqpslevnsarslksvllppsgtqgofltygkirhrktrrgen 671
 Qy 814 KVLNMLNATTTSVLLNLTGAVYSVRLNSPTKAGDQSPSKLSMDPTHEVHPRA 873
 Db 672 e-----tlepnlwlyltglekgeqysfveamvngtgp----- 706
 Qy 874 HPGSTHGRHREGDQLYTHNNNIP--PCD--INPTTHKTKTDYLSGFWLWMLCVILVL 929
 Db 707 -pswvtaetpendi---desqvdqgsslhvrgtn-----clim--- 743
 Qy 930 VISAIAINMYFKRKHQKTELGHLSVSDNEITALINS--KESLWDHRRWRTATDK 987
 Db 744 -----swtppi-npvlrvrglysgyvgyspvaetrzvdskryyslerle 787
 Qy 988 DSGLSKESKLLSVNNSQWYNSDQGT--YAEVTRNLTTFTYCRKSPNPPTPATMII 1046
 Db 788 ss-----shyvislkafnagegvpelyesatrsit-----dptdwyd----- 826
 Qy 1047 GTSSSETCTKTTISADSKDQTSRPSYDAFAGQVAV-----PV-VKSMTLQTPVEPI 1099
 Db 827 -----ypllddftsvpdlstpmppvgvqvalthdavrzs 863

Qy 1100 WSEFLPPPHPPPPSTVGYAQGPSSSKSSKASGISTNWSILMASTSSSSGGFSA 1159
 Db 864 wadsvpkqktsevirlytvrvtsfssakys-----edttislyta 907
 Qy 1160 WGVSP-----QYAVACPPENVISNPLSAGVAGSTONRYOITPIN 1197
 Db 908 tglkptmtyesvnmv-tkarrstsvmtahat--tyeaspts 946

RESULT 15
 W74152
 ID W74152 standard; Protein; 1257 AA.
 XX
 AC W74152;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Human L1 cell adhesion molecule.
 KW
 XX Human L1 cell adhesion molecule; LICAM; neurite growth;
 KW nervous system development; nerve regeneration;
 KW neuronal cell cohesive interaction.
 XX
 OS Homo sapiens.
 XX
 PW US587225-A.
 XX
 PD 16-FEB-1999.
 XX
 PF 18-NOV-1994; 9405-0341843.
 XX
 PR 26-JUN-1992; 9208-0904991.
 XX
 PR 18-NOV-1994; 9405-0341843.
 PA (UYCA) UNIV CASE WESTERN RESERVE.
 XX
 PI Lemmon V;
 XX
 DR NPI; 1999-166719/14.
 DR N-PSDB; X01598.
 XX
 PT Human L1 cell adhesion molecule - supports neurite outgrowth and is
 PT involved in nervous system development and repair
 XX
 PS Claim 1; Fig 3; 45pp; English.
 XX
 CC This sequence is the human L1 cell adhesion molecule (LICAM) of the
 CC invention. LICAM supports growth of neurites in vitro and is involved in
 CC development of the human nervous system and in nerve regeneration. It is
 CC useful in vivo and in vitro experiments on nerve growth and
 CC regeneration. LICAM mediates cohesive interactions of neuronal cells to
 CC each other and to extracellular matrix.
 XX
 SQ Sequence 1257 AA;

Query Match 8.5%; Score 633.5; DB 20; Length 1257;
 Best Local Similarity 23.7%; Pred. No. 3e-28;
 Matches 243; Conservative 142; Mismatches 395; Indels 245; Gaps 36;

Qy 35 WLLVLVNASLNPVAVGQYQSFRILE-----HPTDVLVKNPEATLCKVEGKEPTI 87
 Db 9 wpll-lcspclliqipeeghnmvpeiqtegsprlrvfptddisilcoasgkpevqf 67
 Qy 88 ENFKDGEVSTVNEKSHRVQKDALFPYRMQGGK-----PDQGEWCVAKNRVQVARS 143
 Db 68 rtrdrgvhlkpeelgvtvygsphegsf--titgnnsaqrqfgyrlcrfasknlgstms 125
 Qy 144 RHASLIQVLAEDDRVPEPKDT-----RVAKGETALLECEGPKGPIPEPTLWIKDGVLDL 199
 Db 126 h-----eirlmeagapkvketykpreveegsvlpcnpppsaepliyms----- 174
 Qy 200 KAMSGASSRVIVDGNLLISNV-----EPI----- 226

Db 175 kllhkgdervtmgggnlyfanvltsdnhdyichahfgpttrtiikepidlrkatns 234
Qy 227 ----- 226
Db 235 midrkrllfptnssshlvaqgplvleciaegfptptikwlrpsgmpadrvtynhn 294
Qy 227 -----DEGNKACIAQNLVGTRESSYAKLIQVQKPYFMKEPKQVWMLYQOTATPH 275
Db 295 ktllqkvgeedqeyrclealsgsarhay-yvtvaapywllhkpqshlypggetarld 353
Qy 276 CSVGGDPPPKVLMKKEBGNIPVSARILHDEK-----SLEISNITPTDGETTYCEARN 329
Db 354 cqvqgrpqpvtwr--lnglpre--elakdqkyrigrgaililsvgsdntwtqoearnr 409
Qy 330 WQGISARASI-IVHAPPNPTKRPSNKKVGLNG-VVOLPCMASGNPPSVFVTKSGVSTM 387
Db 410 hglilanaylyvqlpakiltadnqymavqgstayllckafgaprvsgwldedgttlv 469
388 FPNSSHGROQTVAAADGTLIQITDVRQEDEGYVCSAFSVSDSTVRVFLQSVSDVERPPIII 447
Db 470 ----qderffyangtligirdiqandtgryfclaadqnnvntmanlkvdatq-----i 520
Qy 448 QIGPANQTLPGSVATLPCRAICNPS--PRIKWFDHCHAVQA---GNRYSIQSSSLRD 502
Db 521 tqgsrtiekkgsvrtftcqsafdsqpsltwrgdgrldgelgdsdky-fiedgrlvih 579
Qy 503 DLQLSDSGTYTCTASGERGET-SWAATLIVEKPGSTSLHRAADPSTYPAPGTFKVLN-V 560
Db 580 eldysdqngyscavsteldvvearaqllvvgspg-----pvrplvlsdlhl 626
Qy 561 SRTSISLWAKSQBKPGAVGLIGYTYETFSPL-QTGWVAAHRVGDQTVTISGLTPT 619
Db 627 tqsgvrvswspaedhn---aplekydiefedkemapekwyalskvpgnqtsttlklspyv 683
Qy 620 SYVFLVRAEDTQGISVPSGLSNVKTIEADFDASANDL-----SABRTLLTKGSVEL 672
Db 684 hytfvrtainkygpgspvsetvvtpea---apeknvdkvkgegnettnamvltwkplrw 740
Qy 673 IDASAINASAVLENNLHVSADEKHYEGLRIHYKDAVPSAQYHSITVMDASAESPVGN 732
Db 741 ndwnapqvq-yrvqwr-----pgytrgwgqev-----sdqflvsn 777
Qy 733 LKXITYEFLPPFETIRQOPNSKALTU----EDVPSAPPONQIQGNVQTAGVW 788
Db 778 tstfvyvyeikv-----qavnsqgkspqyvtigygedypgalpelegiellnsavikw 833
789 TPPPQHHNGLYKIEVSAGNMKVLA-----NMTLNTTISVLLNMLTGAIVISV 841
Db 834 rpvdlagvghlrygnvtywregsqrhskrhkhdkhvvpantstvislgrlpsayhl 893
Qy 842 RLNSFTKACDGPYSKPSISLFMDPTH-VVHPPRAH-----PSGTHDGRHEGQ 886
Db 894 evqafngrysgspase--fsttpegyvghpealhcsqntallrwwpplshngvity 951
Qy 887 DLTYH 891
Db 952 visyh 956

Search completed: January 22, 2001, 12:17:03

Job time: 1580 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:23:24 ; Search time 325.28 Seconds
(without alignments)
291.200 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPMPENHAIARSTTTNN.....SCLYAEAGEPAPRMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1609.5	21.7	1273	2 T42405	sax-3 protein - Ca
2	1607.5	21.6	1612	2 T30805	duttl protein - mo
3	1585	21.3	1651	2 T14180	transmembrane rece
4	1395.5	18.8	1344	2 T14316	rig-1 protein - mo
5	826.5	11.1	423	2 T29549	hypothetical prote
6	790	10.6	874	2 T29548	hypothetical prote
7	686	9.2	1277	2 T30532	neural cell adhesi
8	677	9.1	1443	2 T50600	neogenin - chicken
9	661	8.9	1896	2 T08851	Down syndrome cel
10	558	8.9	1040	2 A34695	axonal glycoprotei
11	647	8.7	1028	2 T58164	BIG-1 protein - ra
12	645	8.7	1028	2 A53449	plasmacytoma-associ
13	644.5	8.7	1040	2 A94936	transient axonal g
14	644	8.7	1272	2 S26180	neurofascin - chic
15	641.5	8.6	1260	1 S05479	neural cell adhesi
16	637.5	8.6	1447	2 A54100	tumor suppressor p
17	636	8.6	2222	2 T13924	sdh protein - frui
18	633.5	8.5	1257	1 A41060	neural cell adhesi
19	632	8.5	1259	2 A43425	Bravo/Not-CAM cell
20	627	8.4	1259	2 S36126	neural cell adhesi
21	626.5	8.4	1036	2 S22383	axinin 1 precursor
22	624	8.4	1268	1 A39640	neural cell adhesi
23	622.5	8.4	1427	2 T51669	tumor suppressor -
24	619.5	8.3	1018	2 J04211	neural adhesion pr
25	614.5	8.3	1018	2 A54744	contactin 1 precu
26	610.5	8.2	1375	2 T13822	frazzled gene prot
27	604	8.1	1239	1 A32579	neuroglian - fruit
28	602.5	8.1	1020	2 S05944	neural cell surf
29	600.5	8.1	1021	2 A57112	contactin precu

30	595.5	8.0	1010	2 J00094	Fil protein precu
31	595.5	8.0	1091	2 S01998	contactin precu
32	595.5	8.0	1232	2 T43027	neural cell adhesi
33	583	7.8	1907	2 S50893	protein-tyrosine-p
34	568	7.6	1197	2 T30581	neural cell adhesi
35	564.5	7.6	1898	2 S46216	leukocyte antigen-
36	561	7.6	4391	2 A38096	perlecan precursor
37	559	7.5	1209	2 T42718	probable neural ce
38	555.5	7.5	1526	2 T13823	frazzled gene prot
39	547	7.4	1863	2 S46217	protein-tyrosine-p
40	546	7.4	1897	1 TDHUKL	leukocyte antigen-
41	543	7.3	1912	2 A56178	protein-tyrosine-p
42	534	7.2	7962	2 T38346	elastic titin - hu
43	531.5	7.2	1501	2 T58148	protein-tyrosine-p
44	530.5	7.1	1256	2 T03096	CD0 protein - rat
45	530.5	7.1	2029	1 TDFFLK	protein-tyrosine-p

ALIGNMENTS

RESULT 1

T42405

sax-3 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000

C:Accession: T42405

R:Zallen, J.A., Yi, B.A., Bargmann, C.I.

Cell 92, 217-227, 1998

A>Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple

A:Reference number: Z22160; MIM:96117250

A:Accession: T42405

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1273 <I>A>

A:Cross-references: EMBL:AF041053; NID:g2804779; PIDN:AAC38848.1; PID:g2804760

C:Genetics:

A:Note: sax-3

C:Function:

A:Description: sax-3 function is required at the time of axon guidance

Query Match	21.74;	Score 1609.5;	DB 2;	Length 1273;	
Best Local Similarity	31.64;	Pred. No. 7.2e-74;			
Matches	422;	Conservative 195;	Mismatches 526;	Indels 191;	Gaps 39;
Qy	55	SPRIIEHTDVLVKKNEPATLCKYCKRPEPT-IEWFKDGSPVSTNEKK--SHRWQKDG 111	: : : : : : :		
Db	30	APVIEHRIDVYVSGSPATLMC--GAKPSTAKITWYKDGQPVITNEQVNSHRIYVLTG 87	: : : : : : :		
Qy	112	ALFFPYTGMQGS--KEQDGGYCVKANKVQVQSRHSLQIAVLDRDFRVEPKDTRVAKG 169	: : : : : : : : : :		
Db	88	SLELLKVNNGSGKDGAGATCYVASNEBGEKVSNGSKLAMLREDFPRVPRVTQALGG 147	: : : : : : : : : :		
Qy	170	ETALCECPGPGIPEPTLWIKDGVPLDILKAMSPAGSSRRVIVDGNLLINSPIDIS 229	: : : : : : : : : :		
Db	148	EMALVCECPGPGFPFVSVWRKDD-----KELRIQDMPTLTASDGNLLIDPVDSIS 201	: : : : : : : : : :		
Qy	230	NYKCAQNLVSTRESSYAKLVQVPRYPMKEPKDQVMLYGTATPHCSGDDPPFKVLAK 289	: : : : : : : : : :		
Db	202	TYGVANVNVGERSVNSPRLSVTFEPKPEKPDMDTVGAVLDFCRVGDQPOQITWK 261	: : : : : : : : : :		
Qy	290	KEENKIPVSRARILHDEKSLSEINPTDSTGYCVBAHNVQGISARASLVHAPNFTK 349	: : : : : : : : : :		
Db	262	RKNEPMPVTRAYIAKDNRLGRLTIERVQSDGEYCYVARNPAGTLASAGHLRWQAPSPOT 321	: : : : : : : : : :		
Qy	350	RPSNKVGNLVQVLECMASGNPPSPVFWTKRGVSLMFPN--SSHGRQVADGTLGIT 407	: : : : : : : : : :		
Db	322	KPADQSVFAGTACTFETLVQGPSFAYFWKSGEQDQLLPSTVSDGRTVSTPCTGLTIE 381	: : : : : : : : : :		
Qy	408	DVDRDGGTYVCSAFVSDSTVTVFLQVS-----SVDERPPPIQIGPANTLQKGSV 461	: : : : : : : : : :		
Db	382	EVYQVDEGATYCMNAGSSLSKALKVTKVATGNTPAKPPPTIEHGRQMTLVNVS 441	: : : : : : : : : :		

Qy	44	NGLPA-----VRGQYSRIIEHPDVLVKKMEPLTNCKVEKPEPTI	87
		: : : : : : : : : : : : : : : : :	
Db	40	NGTPATPSDDNDSNLGYTSRQLQEDPPPIVHSDPLIVSKGEPMLNCKAEGRPTTI	99
Qy	88	EWFKDGEVPST--NEKSIRVFGKDGALFYFMGGKREQ--DGEWTVCAANNGVGAYS	144
		: : : : : : : : : : : : : : : : :	
Db	100	EWNGKEGVETWDKDPDSRERMLSGSLFLPILVIGRKSSPDGVYICVAANTLVGNVS	159
Qy	145	HASLQIAVLRLDFRVPEKDTNRVAKGETALLCEGPKGPKEPTTILWIWKDGVPLDLAMSF	204

[illegible]

us-09-540-245a-15.rpr

QY	436	V--SVYDERPPIIQIPANQPLKGSVATLPCRTANGSPSRIRKWEHDGAVQA--GNRTY	492
Db	419	IKGASIDQLPPIILQ--GPANQTVILGSSVMPKCRVIGMPQNPQIKDKEDRWLQDQDSQFN	477
QY	493	IIQGSLLVDVLDLSDSGTITCTASGEGRSTMAATLVYKRGPSSTLHRAADPTSTAPP	552
Db	478	LMNGNLIIASIQDMDMGFYSVAKSSIGBATWNSRLKQWDQSGSGPQATPSNPQPP	537
QY	553	GTPKVLNVRSTSLISLWAKSQQKAGVPIIGTVYEFSPDLQTGWIAVHRHQDQTVT	612
Db	538	SQPIVTEYTAISILTN--KPNPQSGATA--TSVIEAFSAAGMTWRTVADGVLETTI	594
QY	613	SLGTFPTSTVTLVRADNTQGISVYPSGLSNVRIEADFDAAASANDLSAALTITKGSVEL	672
Db	595	SLGQPTWTLTLVRAVGWGLSEPSSEPVQTDQSSL--SPRADPKWKGQGLAIVRYM	653
QY	673	IDASAISAVRLKRWHLVSADEKIVVEGLAIHKDASVYQSASQISVIMDASAFVWGL	732
Db	654	EPKTYKTLQNTLQNTLW--VDGQVPLQFGYSNRJGLKQDQSLDQSLDQSPKQSTV	711
QY	733	LKYYTKIEFFLPFFETIEISQSGSKTALYIVDSAPNDQTL--GMYQTAGWKMP	790
Db	712	LPFGAIOIQLVVOVGQSLGAESEPTTASIPSEAPSGPOCAVALGGRNSVSTWMP	771

772 PLPSQGNVITYETQYI-WCLNGSEPHMRASAGVAASVYFSSGLPQIYALVAASATG 830
Db 772 PLPSQGNVITYETQYI-WCLNGSEPHMRASAGVAASVYFSSGLPQIYALVAASATG 830
Qy 851 QYFSPKSLISLMDYTHVHPHPRAHSSTCHDGHGQDLIYHMGNPIDNPITTKTKT 910
Db 831 QGVASAPVLVLQL-PPFAAEFG-----PEVSGLAERLA 863
Qy 911 DYLSPGSLA-----LVCVLITSLASVIAWYFKRKHKTEGLHGLSVSDNETALNIN 967
Db 864 KVLKRPAPLAGSSAACALLGLCPAA-----LYRQKQKLSLTY-ASFATPVAVSF 916
Qy 968 SKSL-----WI-----DHRHG--RTADTDSGLSES 994
Db 917 HSGSLGSSGSSPMGLGPAAPLWASLHPPRPSPSAQPPGSCSPNPDP-DRYNEA 975

976 GISLYLAQTARGANASGEGPVYSTIDPVGEELQTHFGGFPQHSSGDPSTWSQYAPPEWSE 1035

[illegible]

RESULT 5
T29549
hypothetical protein ZK377.3 - *Caenorhabditis elegans*

Mon Jan 22 13:04:18 2001

C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29549
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK377.
A:Reference number: Z20639
A:Accession: T29549
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-423 <KHA>
A:Cross-references: EMBL:U08183; PIDN:AAB52658.1; GSPDB:GN00028; CESP:ZK377.3
A:Experimental source: strain Bristol N2; clone ZK377
C:Genetics:
A:Gene: CESP:ZK377.3
A:Map position: X
A:Introns: 24/1; 142/3; 229/3; 284/2; 408/3

Query Match 11.1%; Score 826.5; DB 2; Length 423;
Best Local Similarity 42.8%; Pred. No. 7.4e-35;
Matches 167; Conservative 61; Mismatches 147; Indels 15; Gaps 6;

Qy	55	SPRIIEHPDIDVWKKMEPAKCNVEKBEPTPE--IENWGDGEPVSTNEKK--SHEVQPKDG	111
Db	29	APVIEHPIDVDVWSGSPATLNC--GAKPSAKITWIKDQGPVITNKQVSHIRVDLG	86
Qy	112	ALFFPYTHMGR--KEQDGEYWCVAIRNQNQVSRSHASLOIAVLDFEYRPEKTPRVAGS	159
Db	87	SLPLLVKSVSGKNDSDAGATVYCASNEHGEVSKNEGSKLAMREDFVWRPRTQALQG	146
Qy	170	ETALCEGPGPKGIEPTLILWIKDQGVPLDLLKAMSFGASSVRIVDGGNLLISNWEIPDG	229
Db	147	EMAVLCSPGKGFPEVPVSKOD---KELRIQDMPRTLLSGHLLIDPVDSDSG	200
Qy	230	NYKICAQNLWVGRBSSTAKLIVQVKPIFMKEKQDVMLYQATFHCVSYGGDPPPKVMLK	289
Db	201	TTQCVANNMVGVSRNPARKLVPEEKPEKDFKDMVDGAAVLDCRVTGCDQPOLITK	260
Qy	290	KEBNIPVSRARILHDEKSLNITSNITPDGCTVCAEHNNGVQISARSLIVHAPNPTK	349
Db	261	RKNEPFTPIYRAIKNGLRIEIVQSDGEGYTCYARNAGTLEASAHVLQVAPFSQT	320
Qy	350	RPSNKVGLNGVQLPCMGASNPSPSVFTMKEGVSLTFMNP--SSHGEGYVAADQTLQIT	407
Db	321	KPADQSPVAGGATATFCTGLVQSPAFWFSKGEQODLLFPSTYSADGRLVSPGTGLTIE	380
Qy	408	DVYREDGEGYTCFSAFVSDSTSVRFVLQVS	437
Db	361	EYRVVDGATYCAVMSASGSKAAIKAT	410

```

RESULT      6
T29548
hypothetical protein ZK377.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #tax_change 18-Feb-2000
C:Accession: T29548
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK377.
A:Reference number: Z20639
A:Accession: T29548
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-874 <N>
A:Cross-references: EMBL:U88183; PIDN:AA852657.1; GSPDB:GM000028; CESP:ZK377.2
A:Experimental source: strain Bristol N2; clone ZK377
C:Genetics:
A:Gene: CESP:ZK377.2
A:Map position: X
A:Introns: 91/2; 356/1; 452/1; 701/3; 746/3; 850/1

```

Query Match 10.6%; Score 790; DB 2; Length 874;
Best Local Similarity 27.3%; Pred. No. 1.3e-32;
Matches 254; Conservative 133; Mismatches 377; Indels 166; Gaps 34;

QY	442	RPPPIIQIGPANTLPGKSVLADPCRATGNSPRKXWGHGHAQ-AGURYSIQGSSRL	500
		: : :	
Db	27	KPPPTIEHGQNQYTLWGSSALPLQCGASGKPPCISLWLRGLDITDSRISQSHSLH	86
QY	501	YDOLLSDSGTYITTCASGERGTSNATVITXKPGVS-LHRAADPSPTYPPAGPTPKVLN	559
		: : :	
Db	87	IADLAKPDGTGYTTCIAKNDGSGTWSASLTVEDHTSNQAQVMPKDPSPFSSPQTPIVN	146
QY	560	VSRITSLSLWAKSGXKPGAVGQITVETVYSPDQGTQWIAAHRGDQVITIGSLPTQ	619
		: : :	
Db	147	VTQTESVLEHW-NAPSTSGAGPTGCIITQYSPDQGTWFENIDPVASVETRIKGLKPS	204
QY	620	SYVLVRAENQGTSPVSGLSNVKLTIEADFAAGN-----DLSAARTLLGKS-YELID	674
		: : :	
	205	SYMFVIREAKGICGTPVSSALVITTSKPAQVSLDKNNKMDIAEKRLSEQLKLEE	264
QY	675	ASAINASIVLSEKMLHVSADKEVYKGLHYH-DASVPSQAHSITMDASGSFVGMH	733
		: : :	
Db	265	VKTIINSTALVYKWKRL-ELSIDITIKWKGPPRTINDQY-VWVTSSTPTENVSLH	320
QY	734	KKTYITZFFLTPF--FETIDSGKSNKALTIEDVPSAPPDKIQGMNTAGVWYWP	790
		: : :	
Db	321	NPTPTIEFVPIPHSGVHSIHGASNSMDVLNRAAPSLPEDVRLRLNLTILRLSWKA	380
QY	791	PPSQHGNLNLGTIKYBSGANTMKVLNMLNATTSTVLLNLTGAVSVRLNLSFKAG	850
		: : :	
	381	PKADGKILGLGFI-VIVGAPNANNNTITNERAASYLHLVGTGHTKRWASNSG	439
QY	851	DGPISTKPSIFMD-PTHHVHPRAHPSGTHGRHGQDLTH--NNGNIPPGDINPTTHK	907
		: : :	
Db	440	VGSASGTSEVMNQDTLEKLA-----AQENESFLYGLINKSHVP-----	480
QY	908	KTTDLSGPMWMLVCLVLLVLSAASIMYFKRKHQNTKELGHLVSVDSEINTALNIN	967
		: : :	
	481	-----VIVLVAILLFVWIIATYVWRNSRNSDGKRSFKINDGYSYMAHN	528
QY	968	SKESLN-----IDHRRGWTADTDKSGLSSEKLLSHVNSQSNYNSD--GGT-----	1011
		: : :	
Db	529	-----LMDVQAQNPQNPMTIATGRMTNNRKGALSLTPQAQDFNCCDITSGTMRPGS	584
QY	1015	---DTABV--DTRNLTPFYNCRKSPDNPPTATYMTIGTSSSECTKITSADQGS	1063
		: : :	
Db	585	EHHYETALQITGGNKNMSTFG-NQYHDDSPATYATTLV-----	622
QY	1068	THSPISDAFAQVPAVPVYSNYLQVPEPINWSFLPPPEHPPTPSSTYTAGSGPSS	1122
		: : :	
Db	623	-----LSWQPA-KWMDKMLRAPAMTN-----VPVPE--PPARYADHTAG--RRS	663
QY	1128	RKSKSAGSGSLTQSLINASTHSSSSGGFSAWQSYQAVACPNPNVSNPLSAVAGGT	1181
		: : :	
Db	664	RRSRAADGRC-----TLAGLELHRTSGGSR-----DSPHTDYSVLQHSDDGT	708
QY	1188	QNRQITPQNHQPOLPAY-PATTPGAGVAPP-NHL-PFATORHAASSTQAGLNARCAQ	1244
		: : :	
Db	709	GSSKERTGERTPFNKLMDFIPPPSPNPPPGGHYDTATRQ-----LWNGSTP	760
QY	1245	SRACNSDALATPSPMQPPPPVPEGWQTVPHPSNHPMTSSNHQITQCSSCSDSHR	1300
		: : :	
Db	761	EDTDS-----YSDGAFARVDNA--RPTSRNRLN-----GGRLKGR	797
QY	1305	SSQSHKROLLEHGSSAKRGGHHRRAP 1334	
		: : :	
Db	798	DDDSORSLMDGGSSEADGENSGDVP 827	

RESULT 7
T30532
neural cell adhesion molecule L1 homolog - Fugu rubripes


```

Qy 679 NASAVLRLHWSADKXVBLRLHYDQSVASQYISIVMDASAESFVGNLKYTX 738
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 745 VITWEMDMKRNKNGPDPFKL-----VMMRVGSSGPDWEEITI-----APPYITVDQWQFSA 797

Qy 739 YEFFLPPFPEIBEQPNSK-----TALYEDVPSAPPNDIQIHYNTAGVWMT 789
      F-----: : : : : : : : : : : : : : : : : : : : : :
Db 798 FE-----IRYQALVWKNGLGPEDPIGTGSEDVPLEALNGLNENSTIRVWS 844

Qy 790 PPPSGHNGNKGYKIEVSAGNTMKVLNMTLN-----ATTSSLVLANLTGAVTSVR 842
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 849 AVOKETVGHGLGTYKTYLTVGHRRNSRQAPENIVMQTGANEXKSLTNLRPCHTDLA 908

Qy 843 LNSFTKAGDGPISKPLSLMDPTHVHPPRA-----HPSGTEDGRHGQDLTYHNNMIPP 898
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 909 ISAFNSKGGFLSEXTS-EMTPEGVPPNMSMWTSPSS-----EITLWHT--PP 956

Qy 899 GDINPTEHKKTDYLSGPMWMLVCIULLVLVISAISVYFKRKHQMTKGLHLSVV-- 956
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 957 SKFN-----GILLGSLQYKRMQSDCNPDQVLDI 985

Qy 957 SDNEITALINIKSLEIWDHGRWR-----TADTKDGLSGESKL-----LSHWNSQSN 1006
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 986 ASPEITLILGL-----LDHSHYQPLLMARTAAGKGLSTIELGATLGLPLPANISLS 1039

Qy 1007 YNNSDGGTDYAEVOTRNKL-----TFYKNRSCDNPTPYATMTIGTS----- 1049
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 -----AERSVNLSEWRKRRHTVQFQHYFSKNGTNGGMMKTEZVNSLSQFF 1089

Qy 1050 -----SSECTKTTISADKSGTSHSPYSDAFQPVPPVVPV 1086
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1090 QLQGLTGPSHRLLPYKNMTWEZELQTGTSVTEVQSPFATQW-----FIGVSAVLL 1146

Qy 1087 -----KSNLYTPVEPINWSEFLPPPE-PPSPSTYGTAGQSPSSRSKSKSA 1134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 LLVLILCPYHSKSGKISVKD-----KEDGPMSEARMPMDTDEYR-SLESDLEEKRT 1202

Qy 1135 -----GSGISTQSLNASHSSSSGSGAWGSPQYVACAPPEN 1174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1203 SQPSLSEKSLCSEDLDFNGSSAVITELMNDSELSAQFSR--HSEGPPEHGV--PDN 1257

Qy 1175 VYSPN 1179

Db 1258 SPLNP 1262

```

RESULT 8
 I50600
 neogenin - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: I50600
 R:Vielmetter, J.; Raynem, J.F.; Roman, J.M.; Dreyer, W.J.
 J. Cell Biol. 127, 2009-2020, 1994
 A:Title: Neogenin, an avian cell surface protein expressed during terminal neu-
 A:Reference number: A55193; MIM:95105243
 A:Accession: I50600
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1443 <VIEW>
 A:Cross-references: EMBL:D07644; NID:a641965; PTDN:AAC59662.1; PTD:a641966

Query Match 9.14; Score 677; DB 2; Length 1443;
Best Local Similarity 23.18; Pred. No. 1.4e-26;
Matches 304; Conservative 171; Mismatches 541; Indels 302; Gaps 49;

Qy 157 FVPEKDFRWAKGETALGEC-----PPKGIPEPFLIWIWGQVLDLKAASFSSGVYR 211
Db 21 FVPEMDLISVRGASVIMNCSSYCTPPK-----ISWKDGT-----LLNLVSDRRQ 68

Qy 212 IVGGNGLILSNW-----EPIDENGYKICAQ-NDVGTPRESSAKLVQVQVPIPKPEKQ 264
Db 212 IVGGNGLILSNW-----EPIDENGYKICAQ-NDVGTPRESSAKLVQVQVPIPKPEKQ 264

```
Qy 1257 PSPMQP-PPVPPVEGQYQVPHSHPMPTSSNKHQYQCSSE-----CS 1300
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1060 PVDGYPDKPLSGNS-----PHGSPISPLDSNMLLWIVSIVGTVITVIVVAVCT 1113
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

Qy 1301 DHSRSSQSHKRLQLBHGSSAKQRG-----HHRR-----RAPVQPCM 1340
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 1114 RRTTSRQKKRAACKSVGSH-KYGNKSKDVKPDLITHERLEKPIDKSPDPNPMT 1170
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
```

T08851

Down syndrome cell adhesion protein 1 - human (fragment)
N:Alternate names: Down syndrome cell adhesion molecule
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: Q08851
R:Yamakawa, K.; Huo, Y.X.; Haendel, M.A.; Hubert, R.; Chen, X.W.; Lyons, G.E.; Korenberg
submitted to the EMBL Data Library, September 1997
A:Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down
A:Reference number: 216495
A:Accession: Q08851
A:Status: preliminary; translated from CB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1896 <YAM>
A:Cross-references: EMBL:AF023449; NID:G169765; PID:G169766
A:Experimental source: brain; developmental stage: 14 weeks; fetal
C:Genetics:
A:Gene: DSCAM
A:Map position: 21q22
A:Note: derived from alternately-spliced mRNA
C:Function:
A:Description: involved in nervous system development
C:Keywords: alternative splicing

Query Match 8.9%; Score 661; DB 2; Length 1896;
Best Local Similarity 24.2%; Pred. No. 1.2e-25;
Matches 305; Conservative 175; Mismatches 473; Indels 308; Gaps 60;

QY	55	SPRIIEHPDIDVLVKNNEPATLCKVEKGPPEPTTIEWFKDGEVPVSNKKSHRQV-----FKDG	111
Db	392	TPKLIISAEKVEVSPAPPEVSLCKNKKGTPLPTITWLLDDPLI--KGGSHRISQMTSBE	449
QY	112	ALFFPTIRMGKKGDDGGEYCWAKNRVQGVASRHASLQIAVLKDDFVPEV-KDTRVAKGE	170
Db	450	NVYSVLTSSSSQVROGQVRCITANMSAG-VWLYQARINV-----RGPASIRPMKNITAJGR	505
QY	171	TALLSGDPKGIPPEPTLKNKGQ--VPDLKAMSGSGSRVRIVDGGNLLISWE-PID	227
Db	506	DTIHC-RVIGTTPYYSIKWYKNSLLFFN-----HRQVAFENNGTKLSDVQKVED	555
QY	228	EGNVCN--IAQLNGLWRESSYAKLVQVKPFI--NKEPDQVMLYGQTATFC-SVGQDP	282
Db	556	EGEYCNVWQQLSTQSQSH--VTVKVPPIIQPEFFRESI--GQRVPICVWVSGLD	610
QY	283	PPKVLKMKKEGNIPVRSARILDD--EKSLEISNITPIDGTYCAHNWQISARSL	339
Db	611	PIITTFWQGRGPIPGSLGVTIONIDFTSSLRISNLSLMHNGNTCIARNEAAVEHQSQL	670
QY	340	IYVHPATNTPKPSNKKVGLNG-VVQLPCMASGNPPSPVFWT-KEGVSTW-PNSSHGKQ	396
Db	671	IYRVPPKFTVPKROD-GYGVKALNCKSABGYPPTIWNKFSKGAGVQPOQPIOTALNGR	729
QY	397	IYAADGLITGLDVRQEDGEGTYTSCAFSVDSSTVR-VQLVYSVDERPPPIIQGPNQOT	455
Db	730	QVLSNGLSLIKVHVEDSYTILCKYVSMDGVGKSSMYLTV-----KIPAMTSPYNTLL	784
QY	456	LPKGSVATLPCRTAGNPSPRIKHWDHCAVQAG-NRYSILIG-----SSLRWDQLDLS	507
Db	785	ATQOGKKEMS-TAGEKPIIVNKEKEDRIINPEMARIVLSTKEVEGEVSIQLQLPTVBE	844
QY	508	DSGYTCTASSEKGETSNAATLVKXPGSLSHRADPSTPAAPGTPLVNLVNSITSISL	567

Db 845 DSGFSCFCHAINSTGDRGLIQLTVQEP-----PDPEI-EIKDKVARTITL 889

Qy 568 RWAKSQKQKPAVPIIGTYVEYFSDLTQGWIAVHRVG-----TQVITISGLTSGTSTV 622

Db 890 RWTMGFD--GNSPITGYDIE--CKNKSDSN--DSAQRTKWSQPLNSATIIDIHPSSTTS 943

Qy 623 FLVRAENTQGISVPSGLSNVIRKTIADFDASANDLSAARTLLTGKSVELIDASAINASA 682

Db 944 IRMANIRKIGSEP--SNEL-TITAD-EAAP-----DGPPE--VLEPESISQS 986

Qy 683 VRLSNMLHVSADERY--VEGLRIHYKDAVSQAQHSITVMDASAES--FVVGNLK 735

Db 987 IRWTW--KAPKHLQGLIRSYQIQREYTSGGNFQNIISVDTSGDSEVYTLIDNLK 1042

Qy 736 YTKYEFPIITFTEIEGQPSNKTALTYEDVPSAPPDNIQIGMYNQAGVWRVTPPPSQH 795

Db 1043 FTQYGLVQACNAGRTGSPSQBIITITLEDVSYPPENVQAATSPESISMSWTLSKEA 1102

Qy 796 RGNLNLGKIEVSAGTKMVLAMLTNATT--SVLNLMLTGAVSYVRLNSFTKAGDG 853

Db 1103 LNLGLQGFV--IYANLMDGLGEKNTITQPSLELDGEKTYNSIQVLAPTRAGDV 1161

Qy 854 YKPSISLNDPETHVPPRAHSCDHRHGBQDLTYHNGNIPPQDINPTKKTKTDYL 913

Db 1162 RSEQI--FTRTKEDVPGP--PAG----- 1180

Qy 914 SGFWMLWCIVLVLVLSAASVYFK-----RKHQMKELGHLVSVDNE-- 960

Db 1181 -----VKAAASASWVPSWLPPLKGLIRKATVTSHPPTVVISSEAS 1226

Qy 961 -----ITALINKESLWDHHRGTADTDKXSGLESKLSVHNSQSYNNSDGG 1013

Db 1227 PDGFSFTRILNRRNQTSVWV-----VATVSAGG-----NSEETI 1262

Qy 1014 T--DTAEVOTRNL-----TPYNCRKSNOPPIATW--IIGTSSTCTKT 1057

Db 1263 TVEPLAKAPARILTFSGVTTPMMDVILPC-KAVGDPSPAVKMDKSNQTPSLVTDGR 1321

Qy 1058 TSISAD-----KDSGTASPSDAFAGQVPAVPSKYNLTQVPEPIMSE-- 1102

Db 1322 KSLFSGSFIRTKAEDSGYS-----CIANN-----NMGSEITL 1358

Qy 1103 --FLPPEPHPP--PPSSTGYAGQSPSSSKSSKAGSGISTYMSILNASBS 1152

Db 1359 NLQVQVPPQPLVTSKTSSTSLTSLWLPD-----NGSGSRGTILQTSEDNS 1407

Qy 1453 SSGGFSAMGVSQYAVACPPENVISNPLSAVAGTQNRQYITPTNRPQPLPATPATGP 1212

Db 1408 EQ--WGSPF--TSPSERSTR--LENLAKGWTYKLTIAQ-----GVGP 1445

1213 G 1213

Db 1446 G 1446

RESULT 10

A34695

axonal glycoprotein TAG-1 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 21-Jan-2000

C:Accession: A34695

R:Furley, A.J.; Morton, S.B.; Manalo, D.; Karageorgos, D.; Dodd, J.; Jessell, T.M.

Cell 61, 157-170, 1990

A:Title: The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neur

A:Reference number: A34695; MUID:90198980

A:Accession: A34695

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1040 <R>

A:Cross-references: GB:M1725; NID:g207148; PID:AAA2201.1; PID:g207149

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

C:Keywords: glycoprotein

F:343-399/Domain: immunoglobulin homology <IMH>

Query Match 8.94; Score 658; DB 2; Length 1040;

Best Local Similarity 24.18; Pred. No. 8.1e-26;

Matches 251; Conservative 145; Mismatches 433; Indels 214; Gaps 36;

Qy 22 SRSSSRMLPAMLLLVLA-----SNGLPAVQGOQSRPIIEHPTLVKK-----NEPAT 74

Db 5 ARKASALLLVLAVALVSPGWSQAQTPATG--PIPEEPGILLPESARDOVT 60

Qy 75 LNCYVEGPEPTIEWFKDGEPVSTNEKSHRVQFKGLAFFTRMGKKRGDGGETMCVA 134

Db 61 LACRASPAPATYKWMNG--TDNLEPGSRHQLMGNLWVI--NSPTKTDQAGTQCL 115

Qy 135 KNRVQAVSRHASLQIANLRDPFVEPKDTRVAKGETALLCGPPEGIPELTIW---- 190

Db 116 SNWPTVYSKLANLRFPLQEGFSKEERDPKTHBGWGLPCNPAPHYGLSLYKLLPE 175

Qy 191 KDGVLDDLKASPGASSVRVIVGGNLLISNVEIDBNYKCIQNLV--GTRE--SSY 246

Db 176 PNPIPTDGRHPS-----QTGNNIARTNASDLNGLSCLATSHDFTKSVSKF 226

Qy 247 AKLIV-----QVKPYMKPEKQVMLVQQTATPHSGVSGPPKVLAKKEGNIP 296

Db 227 AQNLAAEDPLFASIKARF--PPETALVQVQVTECFAPGNPKYKWKVQDGLSL 283

Qy 297 VSRARILDEKSLSEISNITPTDGGTYVCAHNNGVQISARASLIHAPNFTKPSNKKV 356

Db 284 PQWATA--EPTLQISVSPFEDGTYCEANSRGEDTQVGRILVQVQVGLVLSIDETA 340

Qy 357 GLNGVQLPQCHASGPPSVFVTKGVSTLMPNNSHGRQVVAADGLTQITDVRDEGTY 416

Db 341 DIGSNLWCAAGKAPRPMVRLANGE--PLASQNRVLA--GLRFSKLSLEDSGM 394

Qy 417 TVCSASVVD--STVRFVLQVSDVERPPPIIQLGAPQLPAGSVATLPCRAENPSP 474

Db 395 TQCVANRKGITYASAEALVQALAPDPNPNVRLIPA--ARGESILCLQPAKFA 450

Qy 475 RIKWFDHGAQVAGNRYISIQSSRLVDDQLSDSGSTYCTASGERGETSWAATLVE-- 532

Db 451 TILMSKEITGLNSTRVTVSDGTLIIRNISRDEGKYCFAPNFMGAANSTGLSVERDA 510

Qy 533 ----KPGSTSL-----HRAADPS--FY----- 548

Db 511 TKITLAPSADINWGNLTLQCHASHDPTMLPTLTDLPIDFDPKGGHTRASAKET 570

Qy 549 -----PAPGTPKVLNVSRTISLNL 569

Db 571 IGLTLIMAHYHGGKTYCMAQTVVDGTSKENTVLVYRGPQGPVQVQVIGDTTQLSN 630

Qy 570 AKSQKPAVQVPIIGTYVEYFSDLTQGWIAVHRV--GDTVTV--ISGLTSGTSTVFLV 625

Db 631 SRGPDNH--SPIAKTYLQARTPPSGKWKQVNPVNIENGAETAYQLVGLMFMOTFRV 687

Qy 626 RAENTQGISVPSGLSNVIRKTIADFDASANDLSAARTLLTGKSVELIDASAINASAVL 685

Db 688 SASNLGTFSGPSSKIRTKA-VPSVAPSGLSGG-----GGAPGLI-----I 731

Qy 686 ENMLHVSADERYVEGLRIHYKDAVSQAQHSITVMDASAESFVVG--LKRTYEFYFL 743

Db 732 NN-TPVSRQNGDGGYLLSPFRQSSSKQVAVPGADQYFYVGNDSIQTPFFYFKI 790

Qy 744 TPFTEIEGQPSNKTALTY--EDVPAPPDNIQIGMYNQAGVWRVTPPPSQHNGNLY 801

Db 791 RST--NRGQGPESLVALYSAREEPAPAKVAKWAGSSSESMVSN--EPLVDQMGILL 847

Qy 802 GYKIEV-SAGNTMKVLAMLTNATT--SVLNLMLTGAVSYVRLNSFTKAGDGYPSKLSL 850

Db 848 GYIRIKWAGINAAADRVRTAGLTDSRVYGLNPNTKHYTVVTRATRGATGSPASADA 907

Qy 861 FMPTTHVPPRAHPSGT-----HDGRHEGLDQTYHNNKNIPG 899

[illegible]

RESULT 13
 A49356
 transient axonal glycoprotein TAG-1 precursor · human
 N;Alternate names: axonin-1
 S;Species: Homo sapiens (man)
 D;Date: 20-Feb-1995 #sequence_revision 23-Mar-1995 #text_change 24-Sep-1999
 C;Accession: S35508; S28830; A49356
 R;Hasler, T.
 submitted to the EMBL Data Library, September 1992
 A;Reference number: S35508
 A;Accession: S35508
 A;Molecule type: mRNA
 A;Residues: 1-1040 <HAS>
 A;Cross-references: EMBL:X68274; NID:G36574; PIDN:CAA48335.1; PID:G36575
 R;Hasler, T.H.; Rader, C.; Stoeckli, E.T.; Zuellig, R.A.; Sonderegger, P.
 Eur. J. Biochem. 211, 329-339, 1993
 A;Title: cDNA cloning, structural features, and eucaryotic expression of human TAG-1/axonin-1
 A;Reference number: S28830; MUID:93145965
 A;Accession: S28830
 A;Molecule type: mRNA
 A;Residues: 1-296, 'T', 298-1040 <R2>
 A;Cross-references: EMBL:X68274
 R;Tsiotra, P.C.; Karagogeos, D.; Theodorakis, K.; Michaelidis, T.M.; Modi, W.S.; Furley, A.J.
 Genomics 18, 562-567, 1993
 A;Title: Isolation of the cDNA and chromosomal localization of the gene (TAX1) encoding
 A;Reference number: A49356; MUID:94140354
 A;Accession: A49356

[illegible]

[illegible]

F:279-336/Domain: immunoglobulin homology <IMM>

[illegible]

```

QY 679 HDGHEGQDLTHYNNNGINPGDIN-----PTTHKTT-----910
Db 1013 DPISR--TFFTLAARTQVCGEVIVESPALLNEATPTASTWLPPTTELTPAATIAIT 1070
QY 679 HDGHEGQDLTHYNNNGINPGDIN-----PTTHKTT-----910
Db 1071 -----TTTATPTTETPTPELTPTAITPTT-----TATAASTVSTTTTAAERAAATTK 1122
QY 911 -----DTLSGPMVLVLC-IVLLVLVISAASIMVFKRKQMTKELGSLVSDNE 960
Db 1123 ELAYTKHNDVIATQCGWIGLMAKALLVLL--LIVCFIKRS-----GGKYVPRNKK 1173
QY 961 ITALNINSKESLWIDHGRWRATDCKDS--GLSEKSLSHVNSQSNVNSGDDTDA 1017
Db 1174 DEHLNPKWKN--EDGSDYFRLSESDENPLNNSQTSLDGTIKQESD----DSLVDIG 1227
QY 1018 E 1018
Db 1228 E 1228

```

Job time: 1930 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:26:30 ; Search time 162.41 Seconds
(without alignments)
277.386 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPWHPENHAIARSTSTNN.....SCLYAEAGEPAPROMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Aligned: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
1	759.5	10.2	1377	M	NEOL_RAT	P97603	rattus norv
2	731.5	9.8	1493	M	NEOL_MOUSE	P97798	mus musculus
3	710.5	9.6	1461	I	NEOL_HUMAN	Q92859	homo sapien
4	677	9.1	1443	I	NEOL_CHICK	Q90610	gallus gall
5	661	8.9	2012	I	DECA_HUMAN	O60469	homo sapien
6	658	8.9	1040	I	AXOL_RAT	P22063	rattus norv
7	644.5	8.7	1040	I	AXOL_HUMAN	Q02246	homo sapien
8	641.5	8.6	1260	I	CAML_MOUSE	P11627	mus musculus
9	639.5	8.6	1447	I	DCC_MOUSE	P70211	mus musculus
10	637.5	8.6	1447	I	DCC_HUMAN	P43146	homo sapien
11	633.5	8.5	1257	I	CAML_HUMAN	P32004	homo sapien
12	627	8.4	1259	I	CAML_RAT	Q05695	rattus norv
13	626.5	8.4	1036	I	AXOL_CHICK	P28685	gallus gall
14	624	8.4	1284	I	NRCA_CHICK	P35331	gallus gall
15	614.5	8.3	1018	I	CONT_HUMAN	Q12860	homo sapien
16	611	8.2	1239	I	NRG_DROME	P20241	drosophila
17	602.5	8.1	1020	I	CONT_MOUSE	P12960	mus musculus
18	595.5	8.0	1010	I	CONT_CHICK	P14781	gallus gall
19	547	7.4	4393	I	PGBM_HUMAN	P98160	homo sapien
20	546	7.4	1897	I	PTPF_HUMAN	P10586	homo sapien
21	543	7.3	1912	I	PTPD_HUMAN	P23468	homo sapien
22	540.5	7.3	1266	I	NGCA_CHICK	Q03696	gallus gall
23	530.5	7.1	2029	I	LAR_DROME	P16621	drosophila
24	514.5	6.9	1091	I	NCAI_CHICK	P13590	gallus gall
25	499	6.7	1115	I	NCAI_MOUSE	P13595	mus musculus
26	497	6.7	3707	I	PGBM_MOUSE	Q05793	mus musculus
27	480.5	6.5	1070	I	PTK7_HUMAN	Q13308	homo sapien
28	460	6.2	1088	I	NCAI_XENLA	P16170	xenopus lae
29	456	6.1	1051	I	PTK7_CHICK	Q91048	gallus gall
30	449.5	6.1	858	I	NCAI_RAT	P13596	rattus norv
31	449	6.0	853	I	NCAI_BOVIN	P18186	bos taurus
32	448	6.0	1092	I	NCA2_XENLA	P36335	xenopus lae
33	439	5.9	848	I	NCAI_HUMAN	P13591	homo sapien

34	437	5.9	837	I	NCM2_MOUSE	O35136	mus musculus
35	431.5	5.8	761	I	NCA2_HUMAN	P13592	homo sapien
36	431	5.8	725	I	NCA2_MOUSE	P13594	mus musculus
37	415	5.6	837	I	NCM2_HUMAN	O15394	homo sapien
38	414.5	5.6	1913	I	KML5_HUMAN	Q15746	homo sapien
39	396.5	5.3	1906	I	KML5_CHICK	P17799	gallus gall
40	355	4.8	811	I	FS22_DROME	P34083	drosophila
41	353.5	4.8	873	I	FS21_DROME	P34082	drosophila
42	349	4.7	2481	I	UN52_CAEEL	Q08561	caenorhabdi
43	323	4.3	898	I	FAS2_SCHAM	P22548	schistosom
44	314	4.2	1367	I	VGR2_MOUSE	P35918	mus musculus
45	314	4.2	1694	I	SN_MOUSE	Q62238	mus musculus

ALIGNMENTS

RESULT 1
NEOL_RAT
ID NEOL_RAT STANDARD; PRT; 1377 AA.
AC P97603;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR (FRAGMENT).
GN NEOL OR NGN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97015074; PubMed=8861902;
RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-Y.,
RA Ciolotti J.G., Tessier-Lavigne M.;
RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
RL Cell 87:175-185(1996).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, 20
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR ENBL: U68726; AAB41100.1; -.
DR HSP: P56276; 11LK.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; Ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal.
FT NON_TER 1 1
FT SIGNAL <1 2 POTENTIAL.
FT CHAIN 3 1377 NEOGENIN.
FT DOMAIN 3 1074 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1075 1095 POTENTIAL.
FT DOMAIN 1096 1377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 36 105 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 135 197 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 232 296 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 324 386 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 405 502 FIBRONECTIN TYPE-III.

Db 64 FYFLVEPVDTLVRGSSVILNCS-AYSEPSPNIEWKKDGT-----FLNLESDDRRQLLP 116

Qy	215	GNLLISLV-----EPIDEGNYCIA-----QNLVSTRESSAKVLIVKQVYFMKKEQDQW	266
Db	117	DGSLFISNVWBSKHKP-DGSPYQCAVTVNL-GTVISRTAKLVNAGLPAFTSQPESSV	174
Qy	267	LQGTATPHCSVGSDGPPVLVKKKEGNIYPSRAHLDEX-----SLEISMTPTDE	319
Db	175	YVGNISALNCKVWADLVPPVRWEQ-----NQRLPDLDRVLVLSGTLVISMATGSD	227
Qy	320	GTZYCEAHN-WGQISARASLVIVHAPN-----ETKPSM--KKVLGNVQLVLCMAGS	370
Db	228	GLYKIVBSGZGPPKPSDEAEVLVQDEEVLVILVMPFSSMOKVQGSAV--LPCVWSQ	285
Qy	371	NPPSPVWKKGVSTLMTFNSHSGQVVAADGLITDVRDEBGVTVCSAFVSQDSTV	424
Db	286	LPAPVYRMK-----NEEVLDTESSGLVLLAGSGLEISDVTEDDAGTFFC-----IADNKG	338
Qy	431	RVLQVSSVDERRPPIQIGPAQLVPGSAVLPCBATGNSPRFKNHEHCHGAQAGR	490
Db	339	TVMAQELVIVVQPFGLK-QPANIYAESNDIYFCEVCGKPTPIYKWKMGDVIPISDN	397
Qy	491	YSIQGSSLVWDGLLSDSTYCTTASERGETSWAALLV-----EKPGST	537
Db	398	FKIVKEHQLVGLVSSGFGTQCIANDYVNGAQAQGLILEHDVAIPLPPTLSLAT	457
Qy	538	SLHRAADSPPTAPPCPTKVLNVKST-----ISLRWAKSQEKGAVGPIGTGVTSFDL	594
Db	458	TDHLA--PATGFLPSAPDQVAGSLVSTRFKTLVWRTPASDF--GDLTVSYTKVEG	513
Qy	595	QTGIVAAHRHQDTQTVTISGLTPGTSVFLVRAENTQISVPSGLSNVKTIZADFDAS	654
Db	514	DRREVNTSQPGEWQVITQLMPAVTYIFKVMQKHQ-----SGESASPLRVE-----	562
Qy	655	ANDLSAARTLLTKSVLE-----IDASAINAVRLEKMLVHSADEKTEGLRIHYKDA	708
Db	563	-----TQPEVLPGPAPNIRATAPSTSVTWTZPLSGNKE-IQWKLITYMEK	610
Qy	709	SVPSAQHYSTVMDASAEFSVGNLKKTYEFPFLPFETIIBQPSNKTALZIEDVPS	768
Db	611	GTDEKQ-----DIDVSSHSTINGLKKTYESFRVATNKHGGSQDQAVRTLSDVPS	665
Qy	769	APPDNQIGMNVAGVWVRTPPFSQHNGNLGYKIVSAGNTHKVLNATLNAITTSV	828
Db	666	AAPDNISLEVNSKSLVIMQPPSTTQNGQITKRTKRAKSKSDVTEVLVGTQLSQ	725
Qy	829	LLNLLITGAVTSVRLASPTKASGDSPYSKPS--LMDPTHHVHPHRAAPSCTHGRBQG	886
Db	726	LIEGLDGTETINFAVLAIVTQSPATMDLSAETFSLEDTAPVEP-PSLH-----	777
Qy	887	DLTYENNNGIPPDINPTHTKTIYTLGSLPMLVLCVLLVILVISAASMYFRRKHQ	946
Db	778	-----VBP-----LVTISVSWTPEPNQNI	798
Qy	947	TK--ELGLSVYSDEEITALNINSKESLWIDHHRGWATADTKDGLSKSSKLLVSNWSQ	1004
Db	799	VRGYAIGT-----GIGSPAQIIVKDKQRYTITLNDLPS-----SHVYVL	840
Qy	1005	SNYNMSDG-----TDYAEVDRNLITFYNCRKSPDNPTPYATTMI--IGTSS	1050
Db	841	KAFNVWGGIQLYESATVTRPHDTSEVDFLV-----NAPTYVPDPTPMHPVQQA	893
Qy	1051	SETCTKTSIS-AKDDSGTSHSPYSDA-----FAQGVPAVPVYS-----NYLQVPEP	1097
Db	894	SILSDTIRITWADNLSLKKQKITDSRYTVRWKNTIPANTKYKNNANATLSYLVGLK	953
Qy	1098	INWSE-----LPPPEHPPPSSTGYAQGPSES-----SR	1128
Db	954	NLYLESFVMYKGRSSTWMTAGATFVPTSPKQDVTWSEKCKPRTLIIVMQWPSSE	1013
Qy	1129	KSKSAGSGCI--STNGSILNASIHSSSGSGAWGVSP-----QTAWA	1165
Db	1014	ANGKITGVIIYYSTD--VNAELPD-----WYIEPVGNRLTHOIELDTPITYFK	1000

```

DR INTERPRO: IP0001777; -.
DR INTERPRO: IP0003006; -.
DR PFM: PFM00041; fn3; 6.
DR PFM: PFM00047; lg; 4.
DR PRINTS: PR00014; FNTYPEIII.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL      1:      33      POTENTIAL.
FT CBAIN       34: 1461      NEEGGIN.
FT DOMAIN      34: 1105      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM    1106: 1126      POTENTIAL.
FT DOMAIN      1127: 1461      CYTOPLASMIC (POTENTIAL).
FT DOMAIN      67: 136       IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      166: 228      IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      263: 327      IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      355: 417      IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      436: 533      FIBRONECTIN TYPE-III.
FT DOMAIN      536: 629      FIBRONECTIN TYPE-III.
FT DOMAIN      630: 729      FIBRONECTIN TYPE-III.
FT DOMAIN      735: 829      FIBRONECTIN TYPE-III.
FT DOMAIN      850: 950      FIBRONECTIN TYPE-III.
FT DOMAIN      951: 1052     FIBRONECTIN TYPE-III.
FT DOMAIN      1118- 1121     POLY-VAL.
FT DISULFID    74: 129      BY SIMILARITY.
FT DISULFID    173: 221     BY SIMILARITY.
FT DISULFID    270: 320     BY SIMILARITY.
FT DISULFID    362: 410     BY SIMILARITY.
FT CARBOHYD    73: 73       N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    210: 210     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    326: 326     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    470: 470     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    489: 489     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    639: 639     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    715: 715     N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    909: 909     N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC    1248: 1300    MISSING (IN ISOFORM 2).
FT CONFLICT    1461: 15958  G -> N (IN REF. 2).
SQ SEQUENCE    1461 AA: 15958 MW: 7AAB897B59635A21 CRC64;

```

```

Query Match          9.64;   Score 710.5;  Db 1;  Length 1461;
Best Local Similarity 22.3%;  Pred. No. 3.1e+8;
Matches 313;  Conservative 184;  Mismatches 564;  Indels 341;  Gaps 48;

QY  157  FVEPEKDVIRVAKGETALLCEGPPKIEPPELLWIKDGGPDLDLKMSFGASSVNRVVDGG  216
Db    55  FLVEPEVDLTVKSSSSWILMS-AYPESSPKIEWKKDT-----FLNLVSDDRQQLLPDG  107

QY  217  NLLISNV-----EPIDGNGKICQA-NLWGTRESSIARLIVQVKPIKPEKDDQWMLY  269
Db   108  SLFISNVGSHKHKP-DEGTQCTVATVESIGLISIRLARIAGLPTVPGTSGPSESVYNG  166

QY  270  QTATFSCVSGDGGPPVKLWKEBGNIPVSARLHDEK-----SLEISNITPTDGTGY  322
Db   167  NGAILICVFNADLVPEVFKW-----NRQPLLDDORVILKPSGLMISVSNATGGGGLY  219

QY  323  VCAEHN-NVQGISARSLIVHAPNN-----FTKRPKSNKVLKNGVQLPCMASGNPPPS  375
Db   220  RCVSESGGPKYKSYDEVELKVLDPDEVISDVLKQPSPLRVYIGQDVLPCVASGLPTT  279

QY  376  FVFKT--EGVSTLMFPNSSHGQIYVAADGTLITDVRQDEGEYVCSFVDSSTYRVF  433
Db   280  IKMKNREALDT-----ESSERILVAGGSLSLISDVEDAGCTFYC-----IADNGNETIE  330

QY  443  LQVSSVDERPPIIQIGPANOITLPGKSVATLPCRATGNPSRIKWFHGRHVAQNGRYSI  493
Db   331  AQAELETVQAQETFLA-QPNTIAYEHSMDIVPECEVIGKQPTPVKWKNGDMVLPDYFKI  389

QY  494  IQGSLVDDQLQSLDSGTTYCTASGERGETSWAATLVKXGPSLISLHRAADPSTPYAPP  553
Db   390  VKEHMLQVLVLKYSDEGFTYCIASNDVNGNAQGAQLLILE-----HAPATGFLPSAPR  443

QV  554  TPVKLVNRSITSLKAWSOEKGANGVPIGTYEYFSPLOLGTWIVAAHRVDGTDVTSI  613

```

Db 444 DVVASLSTFRLKLTWRTPASDPH--GDLNLTYSVPTKGLIARVRVNTSPHGGVTVIQ 501

Qy 614 GLPIGTSITVFLVRAENTQGISVPSGLSVIKITLADPDAASANDSAARLTGLKGSVEL- 672

Db 502 NLMPTATVYIFRMAQKRG--SGSSAPLRV--TQPVQPLQ 539

Qy 673 ----IDASAENSAVLENNLHVSADKTYEGLRIHYKDAVSQAQHSITVMDASAS 727

Db 540 GPNLPRATASAPISITVTWETPVSQNGE-QNKLKLYHEKGTREQ-----DVDVSHS 593

Qy 728 FVVGNNLAKTYEFLPPFFETIEGQSPNSKALTIVEDVSPAPDNQIGMYTAGWR 787

Db 594 YTLNGLAKTYETSFVRVAYNKGPGVSTPDVAVRTLSDVSAAPNQLSLEVRNSKIMH 653

Qy 788 WIFPFSQHGNNLYGKIEVSATMKVLAMNLNATITVLLNMLTIGATVSVLNSPT 847

Db 654 WQAPAPATQNGITGYKIRYKASKRSQVETLVSQTLQSLIEGLDGTETNFVAALT 713

Qy 848 KAGDGSPKSPIS--LFMDPTHVFPRAHPSGTHGRGSGQDLTYNMGNIIPGGINPT 905

Db 714 INGTGPAIOLASITFESDLDETRVPEY-PSLSL-----VTP- 749

Qy 906 HKKTITDYLSPMLVLCVLLVLVLSAISMYFKRKHQMK--ELAGHSVSDNEITA 963

Db 750 -----LVTSIVGNTPEPMQNLVVRGTAIG-----G 776

Qy 964 LNIINSKESLWIDHGRWRTADTKDGLSEKSLKSNVNSQSNVNSDGG----- 1013

Db 777 IGSHPAQTIQVQYKRYTTEINDPS-----SHVYTLKAPNVGSEIPLYESAVTR 828

Qy 1014 ----TIDYAEVTRNLTITFCRKSPPNPTIATYMI--IGTSSSETCKITTSI-ADKSGT 1058

Db 829 PHTDSVEQLFVI-----NAPTYIPVDPTPMMPVGVQASLSHDTITWADNSLPK 881

Qy 1059 HSPYSDA-----FAGQVAPVAVYKS-----NLYQVPEVINKSEF----- 1103

Db 882 HQIKITSRKTYVKKMTWIPANTKYKNANATLLSYLVGLKPNWLYEPSVMYTKGRSSTW 941

Qy 1104 -----LPPPEHPHPPSSITYGAQGSPE-----SRKSKSAGSGSI--STWQSL 1145

Db 942 SMHTAGTIFELVPTSPKDKVTVSKGKPKIIVNMQPSEANGKITGYITTYSTO--V 998

Qy 1146 NASH-----SSSGSGFSAMGVSQIAY-ACPZEN 1174

Db 999 NAEHDVIEPVGNRLTHQIQELTLDPTIYFKIQAQNSKNGGPMNSAVQRTPKAASD 1058

Qy 1175 VYNSPLASAVAGTQNRQITPTNQRPQLPAYPATITGGGAVVPEH----- 1221

Db 1059 KMPNDGAGSGGSGKSLPDLSGSDYKPPMNSGNSPHGSPSPSLDNLVLIIVSGVITV 1118

Qy 1222 -----PFATQHAASETOAGLNAARCAQSRACNSCA-----LATPMPMP- 1263

Db 1119 VVVIIVAFCTRRITSHQK-----KRAKCSVNGSHYKGNKSDVPEPDMIH 1169

Qy 1264 -----PPVPVPGWQVPRPNSHPMPTSSMHOITQCSCESDRSRSQSHQLQL 1315

Db 1170 LELPKIDKSDPPNPMITDPIPRNSQDITPV-----DNMDSNKHQRNRSY 1215

Qy 1316 BEHSGS-----AKRGHRRHRAV-----VQPCMESENENMLAYEQRQITSDCNSR 1365

Db 1216 KGRHSEDSMTLAGRGMRKPMMPFDSQPPQVVISAHITSLNLPARHHRHSSLSAPAR 1275

Qy 1366 EGDTCSCSSEGLYARAGAP 1387

Db 1276 -----SLEY-HPGSPW 1286

RESULT 4
NEOL_CHICK
ID NEOL_CHICK STANDARD; PRT: 1443 AA.
AC Q90610;
DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NROGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Arcohsauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=EMBRYONIC BRAIN;
RX MEDLINE=95105243; PubMed=7806578;
RA Vlietinck J., Roman J.M., Dreyer W.J.;
RT 'Neogenin, an avian cell surface protein expressed during terminal
neural differentiation, is closely related to the human tumor
suppressor molecule deleted in colorectal cancer.';
RL J. Cell Biol. 127:2009-2020(1994).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07644; AAC59662.1; -.
DR HSP: P80362; 1WTL.
DR INTERPRO: IPR001177; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; Irb3; 6.
DR PFAM: PF00047; Ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein.
FT NOME 1 1
FT DOMAIN <1 1090 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1091 1111 POTENTIAL.
FT DOMAIN 1112 1443 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 102 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 132 194 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 229 293 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 321 383 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 422 519 FIBRONECTIN TYPE-III.
FT DOMAIN 522 615 FIBRONECTIN TYPE-III.
FT DOMAIN 616 714 FIBRONECTIN TYPE-III.
FT DOMAIN 720 814 FIBRONECTIN TYPE-III.
FT DOMAIN 835 935 FIBRONECTIN TYPE-III.
FT DOMAIN 936 1037 FIBRONECTIN TYPE-III.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 139 187 BY SIMILARITY.
FT DISULFID 236 286 BY SIMILARITY.
FT DISULFID 328 376 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1443 AA; 158050 MW; 558C6795579C0B2 CRC64;

Query Match 9.1k; Score 677; DB 1; Length 1443;

FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	2012	DOWN SYNDROME CELL ADHESION MOLECULE.
FT	DOMAIN	18	1595	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1596	1616	POTENTIAL.
FT	DOMAIN	1617	2012	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	39	109	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	138	204	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	239	300	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	328	392	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	421	491	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	518	582	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	610	676	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	704	773	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	802	872	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	885	972	FIBRONECTIN TYPE-III.
FT	DOMAIN	984	1076	FIBRONECTIN TYPE-III.
FT	DOMAIN	1088	1177	FIBRONECTIN TYPE-III.
FT	DOMAIN	1189	1273	FIBRONECTIN TYPE-III.
FT	DOMAIN	1300	1366	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	1380	1463	FIBRONECTIN TYPE-III.
FT	DOMAIN	1477	1562	FIBRONECTIN TYPE-III.
FT	DISULFID	46	102	BY SIMILARITY.
FT	DISULFID	145	197	BY SIMILARITY.
FT	DISULFID	246	283	BY SIMILARITY.
FT	DISULFID	335	385	BY SIMILARITY.
FT	DISULFID	428	484	BY SIMILARITY.
FT	DISULFID	525	575	BY SIMILARITY.
FT	DISULFID	617	669	BY SIMILARITY.
FT	DISULFID	711	766	BY SIMILARITY.
FT	DISULFID	809	865	BY SIMILARITY.
FT	DISULFID	1307	1359	BY SIMILARITY.
FT	CARBOHYD	28	28	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VASPLC	1562	1571	NFATL2NDG3 -> KEARCEKES (IN SHORT

```

      (ISOFORM).
PT  VARSPLIC  1572 . 2012      MISSING (IN SHORT ISOFORM).
PT  CONFLICT  1893 . 2012      HRPCDILHLPYLMDFLWLNKGGGTGSHDSLQACLEPK
                                     SRLKRPVLEP;PMKAAASSTBQSGWQGAATILPQ
                                     BGAEILQAAMSSWSSDLSRGLHKNWPTASTYLV ->
PT                                     IGVVSTICLHLEWTFQ (IN REF. 1).
PT                                     .
SQ  SEQUENCE  2012 AA; 222259 MW;  0B33CFB781A0B334 CRC64;

Query Match      8.9%; Score 661; DB 1; Length 2012;
Best Local Similarity 24.2%; Pred. No. 1.4e+25;
Matches 305; Conservative 175; Mismatches 473; Indels 308; Gaps 66;

Qy  55  SPRIIEHPDVLVKKNEPATLCKVKKPEPPIEWFGKGEVPSTNEKSHRWQ----FKDG 111
Db  406  PKIILSAFSEKSVAPAEVPSLCKMCKVGLPFLPTITWTLDDODPLL--KGSISRISQMITSS 453

Qy  112  ALFFTRITMGKKKQDGETYCVAKNRVQGVASRSHASLIQVLRDFRVEP-KDTRVARGE 170
Db  464  NVVSTLNTSSSQVDGGVRCRTANNASG-VLVQARILV---RGPASIRPMKNITAIAGR 519

Qy  171  TALLBGKPPKGIPEPTLIWIKG:-VPLDLDKAMSGASSVRIVPGGNNLISNVE-PID 227
Db  520  DTITBIC-RVIGYPTYSIKWYKSNLLFPN-----HRQVAFENGLTKLSIDVQKEV 569

Qy  228  BGNKTC--IAQNLWGTRESSYAKLIVQKVPYI--MKEPKDQVWLQGTATFHC-VSGVGP 282
Db  570  BGEYTCNVLLWQPLSTSQSBV--VTVKVPFPIQGEFFPRFSI-----QGRVFPCVWVSGOL 624

Qy  283  PKVLMKKKEGNIVPSRARILHD---EKSLEISNITPDGETYCEAHNNVQGISARASL 339
Db  625  PITITWQDGGPIRPGSLGVTINDIPDTSLSRISLNSLHMNGNWTIARNEAAVEYSQGL 684

Qy  340  IVHAPNTTKRPSKNKGLVNG-VVQLPCASGNNPPSPVFVT--KEGVSTLMP-PNSHGRO 396
Db  685  IWRVPKPKVQVPODQD-GIYKAGVILKCSABGYVPTPIVWFKSGAGVQPOPIALNGRI 743

Qy  397  VYAADTQLITVDQRQEDGETYCVSAFVSVDSTVR-VFLQVSYDVRPEPIIQLGFANQT 455
Db  744  QVLSWGLSLIKHWVEDSTGYTLCKVSDNVGADYSKMYLV---KIPAMITSYPTNTL 798

Qy  456  LPKGSVALPCLRAITGSPSPRIKWHEDHSHAVQAG-NRYSITQG-----SSLVDVQLLS 507
Db  799  ATQGGKMKSTAHGSKPIIVWKEKDRILNEMARYIVSTKEWGEVISTLIQLITVRE 858

Qy  508  DSGYTTASISGERGETSNMALTVEKPSGTSLHRAADSTYPAPPOTPKVNLNKSITSL 567
Db  859  DSGFSCBAINSTGEDRGIIQLTVQEP-----PDPEEI-EIKWKAARTITL 903

Qy  568  HWAKSGKQKGVPIIGVITVEYFSPDLQTGWIVAAHRGSD---TQVITSLTPTGSTV 622
Db  904  RNTWGFQ---GNSPITGDIIE-CNKKSDSW-DSKQKDVSPQLNASTIADIHPSSTS 957

Qy  623  FLVRAVNTQGISVPSGLSVWIKTIEADFAASANDLSAARTLLTKGSVELIDASAINASA 682
Db  958  IRMTAKNKISKEP---SNEL-TITAD-EAAP-----DGPQEP-WLEPISSQS 1000

Qy  683  VRLMLNLSVADSKY----VEGLRIHYKDAVSPSAQYHSITVMDASAES--FVWNLKK 735
Db  1001  IRVW---KAPKHLQNLGIIRGQIYQREYSTGNGNPFMIISVDTSGSEVITLONLAK 1056

Qy  736  IYKYEFFPLTFETIEQGPSKNKALTIVEDVSPAPPDNIQIOMYNTAGWNTVPPSGH 795
Db  1057  PTYGLVQQA-NRAGTGSQSSLIITTTLEDVPSYPPENQVIAITSPESITSKNSLSKEA 1116

Qy  796  RGNGLNIGYIKSVAGNMTKVLANTLNATT---SVLLNNLITGAVYSVRLNSITKAGDG 853
Db  1117  LNLIGLIFRY-IWANMLDGLGEIKNITITQPSLELDGLKTYNTSIQVLAFTAGDGV 1175

Qy  854  YKSPILSMDPTHVHPPRAHPSGTHDGRHQDQLTYHNNGNPPGGINPTTHKKTOTYL 913
Db  1176  RSBOI--PTTKEDVPGG--PAG----- 1194

```

Qy 914 SGPMVLMLVCLVLLVLSAALSMTYFK-----RKHQMTKGLHLSVSDNE-- 960
 Db 1195 -----VKAASASNVFVSPMLPLKNGIKRITVPCSHPTPTVISSEFAS 1240
 Qy 961 -----ITALNINKESLWIDHHRGWRATDQKSGSEKLLSHVNSQSYNNSSGG 1013
 Db 1241 PDSFTSRITPLNSRHWQTSVWV-----VAVTSAGG-----NSSIEI 1276
 Qy 1014 T-DYAEVDTRNLT-----TFYNCRSPDNPTPIATM--IICTSSSETCTK 1057
 Db 1277 TVEPLAKAPARILFTSGVTTPFMKEDVLPK-KAVGDPSPAYKWKMSNGTSLVTDGR 1335
 Qy 1058 TSISAD-----KDSGTHSPYSDFAGQVPAVVPVKSILQYVPEPINMSE---- 1102
 Db 1336 KSISFNSGFIITVKAEDSGTYS-----CIANN-----MWSGDEIIL 1372
 Qy 1103 -PLPPPPPEHP-----PPSTGYAGQSPSSRSKSSGSGISTNQSLNISRSS 1152
 Db 1373 NLQVQVPPQRLZYKSTTSSSITLSWLPGQ-----NGGSIIRYLIQSEDS 1421
 Qy 1153 SSGSGFSSQVAVACPPENYVSNPLSAVAGQTNQYQIPTNQSPQPLPAYPATIGP 1212
 Db 1422 EQ-----WSPFP-----ISPSERSTR-LENKCACTWKFPTLTAQM-----GVGP 1459
 Qy 1213 G 1213
 Db 1460 G 1460

RESULT 6

AXOL_RAT
 ID AXOL_RAT STANDARD; PRI: 1040 AA.
 AC P22053;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE AXONIN-1: PROCURSOR (AXONAL GLYCOPROTEIN TAG-1).
 GH TAXI.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 31-41.
 RC TISSUE=SPINAL CORD;
 RX MEDLINE=90199890; PubMed=2817872;
 RA Furley A.J., Morton S.B., Manalo D., Karageorgos D., Dodd J.,
 RA Jessell T.M.;
 RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily
 member with neurite outgrowth-promoting activity".
 RT Cell 61:157-170(1990).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
 AXONS. MAY BE INVOLVED IN CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
 CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
 CC -1- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS, AND IN ADULT
 CC BRAIN, SPINAL CORD AND CEREBELLUM.
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED ON A SUBSET OF AXONS
 CC IN THE DEVELOPING RAT NERVOUS SYSTEM.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M31725; AAA42201.1;
 DR PIR: A34695; A34695.
 DR INTERPRO: IPR001777;
 DR INTERPRO: IPR003006;

DR PFAM: PF00041; fn3: 4.
 DR PFAM: PF00047; lg: 6.
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1.. 30
 FT CHAIN 31.. 71015 AXONIN-1.
 FT PROPEP 71016.. 1040 REMOVED IN MATURE FORM.
 FT DOMAIN 56.. 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150.. 218 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256.. 315 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 343.. 404 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 435.. 497 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 525.. 596 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 608.. 614 GLY/PRO-RICH.
 FT DOMAIN 613.. 708 FIBRONECTIN TYPE-III.
 FT DOMAIN 716.. 811 FIBRONECTIN TYPE-III.
 FT DOMAIN 818.. 910 FIBRONECTIN TYPE-III.
 FT DOMAIN 911.. 1005 FIBRONECTIN TYPE-III.
 FT SITE 796.. 798 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 78.. 78 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 200.. 200 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 206.. 206 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 463.. 463 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 479.. 479 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 500.. 500 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 527.. 527 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 777.. 777 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 832.. 832 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 920.. 920 N-LINKED (GLCNAC..) (POTENTIAL).
 FT CARBOHYD 942.. 942 N-LINKED (GLCNAC..) (POTENTIAL).
 SQ SEQUENCE 1040 AA; 113042 MW; 65707EF651C84FB CRC64;

Query Match 8.9%; Score 658; DB 1; Length 1040;
 Best Local Similarity 24.1%; Pred. No. 8.6e-26;
 Matches 251; Conservative 145; Mismatches 43; Indels 214; Gaps 36;

Qy 22 SRSRSRMLLPAWLLLVLA-----SNGLPVARGQYSPRIEHTPLVYKK---NEPAT 74
 Db 5 ARKASLLLLATVLAIVLSPGNSFAQTGPTG---PIFEQGLGLPFESSEDQVT 60
 Qy 75 LNCYVGKPEPTIEWKDEGPEPTNEKSHRVQKDALFYRTMGKKKGQGGETWQVA 134
 Db 61 LACRAASPPATYRWKMG---TDMNLEPSRHGLMGNVLI---MSPTKTDQAGVQCLA 115
 Qy 135 KNRVQAVSRASLQIAVLRODFRVEPEKDIRVAGETALLCGPKGIPPELTIWL---- 190
 Db 116 SNPQVTSKIAVLRFGLFQESKEEDPVKTHGHWGLMPCNPARYGLSTRMLNFE 175
 Qy 191 KDGVPDLKMGSGASSRVRVDGNLLNSIVPEIDEGNTKICIAQLNV---GTRE--SSY 246
 Db 176 PNPIPTDGRHVS-----QTGNLYIARINAGSLNGLSCLATSHNFTSKVSFSG 226
 Qy 247 AKLIV-----QVKPYFMKEPKDQVLMYQQTATFCHSVGGDPPPKVLKKKEGNEP 296
 Db 227 AQNLNADDEPLFAPSIAKRF---PPELTALVQQLTCEFAFGNPNFIRKRWKVSLS 283
 Qy 297 VSRARILNDEKLSLEINPTDSETYVEAHNNVQGISARASLIHAPNFTKRKNPKV 356
 Db 284 PQWATA---EPTLQIPSVSFEDETCENASRGKRTVQGRILVQAPKWLAVISDTEA 340
 Qy 357 GLNGVQLPMASGNPSPVFFVFTKEGVSTLMPFNSSHGKQVVAADQLTQITDVRGE 416
 Db 341 DISGNLWGCAGKGRPMYRWLRGE---PLASQNVFLVA-GDLRFKSLSLSDSGN 394
 Qy 417 YVCSAFVSDG---STVRVFLQVSDVERPPPIIQIGPANTQLPKGSVAZLPKRAZGNP 474
 Db 395 YQCVANREKGTIVASALQVALADPPQNMVRLPIA---ARGGEISILQCPRAAPKA 450
 Qy 475 RIWFNDRGAVQAGNRTSIIGSGSRVDDQLSDSTYTCTASGERGETSWAATLTV-- 532
 Db 451 TLMSKGTILGNSTRVTVTSDQTLIRNISREDEKTYTCFAENFWKAMSTGILSVDA 510

QY 533 ----KPGSTSL-----HRAADPS--TY----- 548
 Db 511 TRITLAPSDANGVNDLQCHASHDPTMDLTPWTLDDPFDIDRPGGHTYRASAKET 570
 QY 549 -----PAPGPTPKVLNVSRTSISLW 569
 Db 571 IGDLTILNHNVRHGKTYCMAQTVVGDTSKEATVLVRGPGPGGVYVRIGDITVQLSW 630
 QY 570 AKSQKPGKGVPIIGTYEYFSPDLQTCWVAAHVV---GDTQV--ISGLTPGTSYTVLV 625
 Db 631 SRGDFNH--SPIATYLIQARTPPSSKKQVTRPNVPIEGNAETAQVLGLMFWMDYFV 687
 QY 626 RAZNTQGISVPSGLSNVKTIEADFDAASANDISAARTLTKGKVELDASAINASAVRL 685
 Db 688 SASNIGLGTGEPGSPSSKIRTKEA-VPSVAPSGLSGG---GGAPGELL-----I 731
 QY 686 EMLWLASAERTYEGRLHYKDKASVPSAQYHSITVMDASAEFVYGVN--LAKITYEYFL 743
 Db 732 MW-TPVSRQYNGDGGPGLLSFRQSSSQWQARVPGADQYFVGMDSIQPTTFPEVKI 790
 QY 744 TPFETIEGQPSNKSATLY--EDVPSAPPNDIQMGYITAGWVTPPTSPHSHGLY 801
 Db 791 RST---NRGDDQPESTALVYSAEERPVAPAKVNAKSSSSSMVSW-EPVLQDMGILL 847
 QY 802 GYKIEY-SAGNMKVLNMTNATITSVLLNNITGAVSVRLNSPTKAGDGPTSKPISL 860
 Db 848 GYTIKWKAGDNZAAADRVRTAGLDTSAVYGLNPNTKHTVRYAYNRASTGPASPSIDA 907
 QY 861 FMDPTKHHVPPRAHPSGT-----HDGHEGQDLTYHNNGNIPPG 899
 Db 908 MT-----VKPPPRAPGNISWTFSSSLSLWDPVPLVNESTVYCYKMLYQ----- 955
 QY 900 DINTPHTKKTTDVLGSEWMLVLY 922
 Db 956 DLAPT---PTLHLTKSNWIEIPV 975

RESULT 7
 AXOL_HUMAN
 ID AXOL_HUMAN STANDARD; PRT; 1040 AA.
 AC Q0246;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE AXONIN-1 PROCURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL
 GLYCOPROTEIN 1).
 RF TAXI OR TAG1.
 CC Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=BRAIN;
 RX MEDLINE=93145965; PubMed=8425542;
 RA Hasler T.H., Rader C., Stoeckli E.T., Zuellich R.A., Sonderegger P.;
 RT "cDNA cloning, structural features, and eucaryotic expression of
 RT human TAG-1/axonin-1.";
 RL Eur. J. Biochem. 211:329-339(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=BRAIN;
 RX MEDLINE=94140354; PubMed=8307567;
 RA Tsiotra C.P., Kargogeos D., Theodorakis K., Michaelidis M.T.,
 RA Modi W.S., Furley J.A., Jessel M.T., Papanathakis J.;
 RT "Isolation of the cDNA and chromosomal localization of the gene
 RT (TAG1) encoding the human axonal glycoprotein TAG-1.";
 RL Genomics 18:562-567(1993).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
 CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
 CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X68274; CAA04835.1; .
 DR EMBL: X67734; CAA7963.1; .
 DR PIR: S28830; S28830.
 DR MIM: 190197; .
 DR INTERPRO: IP0001777; .
 DR INTERPRO: IP0003006; .
 DR PFAM: PF00041; E3; 4.
 DR PFAM: PF00047; I; 6.
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1..28
 FT CHAIN 29..1012 AXONIN-1.
 FT PROPEP 1013..1040 REMOVED IN MATURE FORM (POTENTIAL).
 FT DOMAIN 54..118 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 148..216 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 254..313 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 341..402 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 433..495 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 523..594 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 606..612 GLY/PRO-RICH.
 FT DOMAIN 611..706 FIBRONECTIN TYPE-III.
 FT DOMAIN 714..809 FIBRONECTIN TYPE-III.
 FT DOMAIN 816..908 FIBRONECTIN TYPE-III.
 FT DOMAIN 917..1003 FIBRONECTIN TYPE-III.
 FT SITE 794..796 CELL ATTACHMENT SITE (BY SIMILARITY).
 FT CARBOHYD 76..76 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 198..198 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 204..204 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 461..461 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 477..477 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 498..498 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 525..525 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 630..830 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 918..918 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 940..940 N-LINKED (GLCNAC...) (POTENTIAL).
 FT LIPID 1012..1012 GPI-ANCHOR (POTENTIAL).
 SQ SEQUENCE 1040 AA; 113393 MW; 25487803C28EPB6 CRC64;

Query Match 8.7%; Score 644.5; DB 1; Length 1040;
 Best Local Similarity 23.7%; Pred. No. 4.1e-25;
 Matches 245; Conservative 149; Mismatches 416; Indels 223; Gaps 39;

QY 36 LLLV---LVASNLGPAVAGQYQS--PRIIEHTDLVVK---NEPATLNCKVGEAPEPT 86
 Db 11 LLLVAAVAVLVSSANSALGSGQTTPGVPDQSLVLPSTESTQGVLLACARASAPAT 70
 QY 87 IDWFKDGLVPSTNEKSHKQVKGALFYFRTMGKKEQDGGQGVKAVNRQVAVSRHA 146
 Db 71 TRKKMGKMTNLEKPSRRH--QVGGNLI---NMPTKADQAGVYQCLASNPVGTIVSREA 125
 QY 147 SLQIAVLRDVPFKVPTKATGKATALLCGPPKPIEPLTWI---KDGVLDDLAAM 202
 Db 126 ILRFGLFQEGSKERDPVKAHGQGVMLPCNPAPGSLSTWLNLEFPNFIPTDGRHFY 185
 QY 203 SPGASRVRIVDGNLLISNVEPIDEGNYKICIAQLV--GTRE--SSYAKLV----- 251
 Db 186: S-----GTTGNLIARTNASLQNYSCLAITSMDSTSVFSFKQALAAEDTRL 236
 QY 252 ---QVAPIMKPEKQVLMYQATPHCSVGGDPPPKVLWKEGPNVSRARILHDEKS 308
 Db 237 PAFSIRAP---PAETALVGQVQLCFAFGNPVRIRKWDGSLSGQWTA---EPT 290
 QY 309 LEISNITPD6GTTCVAAHNVQGISARSLVAPNPTKSPNKKVGLNVGVQLPCA 368

Db 291 LQIPSPVYEDGTYTCBAENSKGRDVGRIIVQAPENLVKVIDEADTGSNLHWGCAA 350
 Qy 369 SGMPFSPVYMEKGVSTLMPNSSHGRQVYAADGLQITDVRQDEGTYVCSFASVDS 428
 Db 351 AGSPRPVYMLNAGE---PLASQNVFLA-GDLRFSLKSLSDGMDTC---VAENK 400
 Qy 429 TVRFVLQVSSVDEPPIIPIQANPQLT---KGSVATLPCRTAGSPRIKVPFHGHAQV 486
 Db 401 HGTVIASAEALQVALPDLRFLPVRRLPARGGELLPCQPAKPAKVLVMSKGTIELV 460
 Qy 487 AGNRYISQISSLAVDQLSDSGTTCYASGERGETSWAATLV----- 531
 Db 461 NSSRVPTVPGTLLIRNISRSDBGKTYCPAENFMKANSTGLISVDRATKITLAPSSADI 520
 Qy 532 -----EKPG----- 535
 Db 521 NLGDNLTLCQASHDPTMDLTPFTWLDPPIDPKPGKHRYRTNVKETIGDLILNAQLR 580
 Qy 536 -----STSLHRAADPSTY-----PAPPGTPKVLNVSRSTSLSLWAKSKPKGAVGP 581
 Db 581 HGGKTYCMQAVVDSASKEALVILVAGPPGPGVVRDIDGTTILQNSGRFQNH---SP 637
 Qy 582 LIQTYTFEYFSP-----DLQGTIVAAHVRGDTQVT-IGSLTPGTSIVFLVAENTQGIS 634
 Db 638 IAKTYLQARTPAGNWKQVRTN---PANIEGNAETAQVGLTFLPMDYEFVIAISNLGTG 694
 Qy 635 VPGSLNWKITIEADPDASANDLSAARTLLTKSVELLDASAINAVRLDMLHVSAD 694
 Db 695 EPSPGSKKITRREA-APSVAFPSGLSG---GGAPEL---VWMTFMSKEQ---WGD 742
 Qy 695 E-KYVEGLR---IHYKDSVPSAQHYSITVMDASASFPVGN--LAKTYKIEFFLTPFP 747
 Db 743 GPGLLLSFGQSGSTWQATKVPG-----ADQAPYFVSNESVAPPTPEVKIKST- 791
 Qy 748 ETIYGSGNSKATLT--EDVPSAPPDQIGMYTAGVWVTPPQSHNGHLYGTIK 805
 Db 792 -NKGQDGSFALVYSAEERPVAPTKWAGVSSSEMMNVW-EPVQDDNNILGLTIE 849
 Qy 806 EV-SAGNTMKVLANMTLNAITSVLLNMLTGATVSVRLNSFTKAGDQSPKSI-PLMD 863
 Db 850 RYKAGDKGEAADRVRTAGLTSARVSGLPNTKYHVTYVATNAGTGPASPNAMTX 909
 Qy 864 PTHRVHPPRRPSPG-----THDRHGQDGLTNNNGNPDPGNDP 903
 Db 910 P-----PPRRPGHISWTFSSSLSIKNDPVVPRNESAIVGYKMLINDLE-----LTP 959
 Qy 904 TTKKTTDILSGP 916
 Db 960 TLLHNGNMWIEP 972

RESULT 8
 CAML_MOUSE
 ID CAML_MOUSE STANDARD; PRT; 1260 AA.
 AC P11627;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN L1CAM OR CAML1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-BRAIN;
 RX MEDLINE=88318924; PubMed=3412448;
 RA Moos M., Tacke R., Scherer H., Teplow D., Frueh K., Schachner M.;
 RT "Neural adhesion molecule L1 as a member of the immunoglobulin
 RT superfamily with binding domains similar to fibronectin.";
 RL Nature 334:701-703(1988).
 CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE

CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X12875; CAA31368.1; -
 DR PIR: S05479; S05479.
 DR HSP: P20241; 1CPB.
 DR MGI: MGI:96721; L1CAM.
 DR INTERPRO: IP0001777; -
 DR INTERPRO: IP0003006; -
 DR PRAM: PF00041; fn3; 4.
 DR PRAM: PF00047; ig; 5.
 DR PRINTS: PR00014; FNTPEIIL.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
 FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1124 1146 POTENTIAL.
 FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 346 410 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 404 503 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 531 599 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 827 896 FIBRONECTIN TYPE-III.
 FT DOMAIN 932 994 FIBRONECTIN TYPE-III.
 FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III.
 FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 824 824 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 875 875 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 968 968 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 978 978 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 1180 1183 MISSING (IN SHORT ISOFORM)
 FT (BY SIMILARITY).
 SQ SEQUENCE 1260 AA; 140968 MW; 22BES7001CB2A538 CRC64;

Query Match 8.64; Score 641.5; DB 1; Length 1260;
 Best Local Similarity 23.04; Pred. No. 7.4e-25;
 Matches 275; Conservative 167; Mismatches 458; Indels 295; Gaps 48;

Qy 35 WLLVLVASNGLPVAVGQYQSRIE-----HPTDLVVKNEPATLNCVKSGKPEPTI 87
 ||| : : : : : ||| : : : : :
 Db 9 WPLL-LCSCLLIQIPDQYKGHVLEPPVITQSGPRKLVFPFDIOLKLCARGRQVFEV 67
 ||| : : : : : ||| : : : : :
 Qy 88 EWFKDGHPVSTNEKKS---HVQFKDGLAFVYRTMGKK---BQDGEYVCVANKVNGQA 141
 ||| : : : : : ||| : : : : :
 Db 68 RWTGDKTHFKPEELGVVHEAPY-SGSP---TIEGNSFAPQGVITYCYANSLGTA 122
 ||| : : : : : ||| : : : : :
 Qy 142 VYSHASLQIAVLRODFRPEVKDT---RVAKGETALLEGPPKGIPEPTLWIKDGV--- 194
 ||| : : : : : ||| : : : : :
 Db 123 MSH---EIQLVAGGAPKWKETKVPVEVEGGSVLPKCPNPSPAPRIYWNKSIKFI 178
 ||| : : : : : ||| : : : : :
 Qy 195 -----PLD-----DLK 200
 ||| : : : : : ||| : : : : :
 Db 179 QKDERVSMQNGOLYFANVLSDNHSDYICNAPGRTITIIQKEPIDLVKPTNSWIDRK 238
 ||| : : : : : ||| : : : : :
 Qy 201 ---AMSGASSRVIVDQGNLLIS----- 221
 ||| : : : : : ||| : : : : :
 Db 239 PLLFFLVNSSLRLVALQGSLLLEICIAEGPTPTIKLWLRSDPMPTDRVIYQNNKTLQL 298
 ||| : : : : : ||| : : : : :
 Qy 222 -NVEPIDEGNYIAQMLVGTRESSAKYLIVQVYFMKEPKDQMLYQGLATFHCSSVG 280
 ||| : : : : : ||| : : : : :
 Db 299 LNWGEDEDDYTCILAENSLGSAHAY-VYTPVAPYWLAKPQSHLYGPGETALRDQWQV 357
 ||| : : : : : ||| : : : : :
 Qy 281 DPPPKVLWKEGNNPYSRARIHDEK-----SLEISNITPTDGTIVCAHNNVQGIS 334
 ||| : : : : : ||| : : : : :
 Db 358 RQPEITWRIIN---GMSMETVNDKQTRIEQSGLLSWQPTDMVTCQBARNGSLLL 413
 ||| : : : : : ||| : : : : :
 Qy 335 ARASL-IVHAPNFTKPSKSKVGLNG-VYQLPCMASGNPPSPVFWTKECVSTLMPNKS 392
 ||| : : : : : ||| : : : : :
 Db 414 ANAYITVQLPARIILTKDNTYVNAVBSATALLKAPGAPVSVQWLEDEGTVL---Q 469
 ||| : : : : : ||| : : : : :
 Qy 393 HGRQYVAADQTLQITDQDEGGYVCASFVSSVTVNVLVQSVDERPPIIQLGPA 452
 ||| : : : : : ||| : : : : :
 Db 470 DERFFPYAMGTLISRLQANDTGRYQCPAANDQNNVITLANQVKEATQ---ITQGR 524
 ||| : : : : : ||| : : : : :
 Qy 453 NQTLPKGSVATLPICRATGNPS--PRIKWFHDGHAQA---GMSYISQSSLRVDQLQS 507
 ||| : : : : : ||| : : : : :
 Db 525 SAIEKKAQVTPYQCAFQSPQLQASITWRGDSGLQERQDSQI--PTEKMLVQLSGLTS 583
 ||| : : : : : ||| : : : : :
 Qy 508 DSGTITCTASGERGT-SWAATLVKPKGSTSLHRAADPTTAPPQTKVLVNSRTSIS 566
 ||| : : : : : ||| : : : : :
 Db 584 DQMSVCASTELDEVSRAQLLVVSGSGPPHLELSRHL-----LAQSQVH 631
 ||| : : : : : ||| : : : : :
 Qy 567 LHWAKSQKPGAVGPIGTYVTFSPDL-QGSHVAAHVRGQDTVTSIGLTPGTSTVFLV 625
 ||| : : : : : ||| : : : : :
 Db 632 LWSGPAEDHN---SPIEKYDIEFEDKNAPEKWSLKVPGNQSTTLKLSPVHTVTFV 688
 ||| : : : : : ||| : : : : :
 Qy 626 RAENTQGISVPSGLSNVITKEADFAASANDLSAA-----RTLLAGKSVELDASAI 678
 ||| : : : : : ||| : : : : :
 Db 689 TAIKLVGSGSPVSVESVTPPEA---APEKNPVVDEGNETNWNKLVKLRQWONNAP 745
 ||| : : : : : ||| : : : : :
 Qy 679 NASAVLEMLNHLVSADEKTYEGLRIHYKDSVYSAQHYSTVMDASASFVGNLKXTY 738
 ||| : : : : : ||| : : : : :
 Db 746 QIQ-YRQWQ-----PGQKETWRQVSDPFLVSNSTVFP 782
 ||| : : : : : ||| : : : : :
 Qy 739 YEFLTPFFETIEGQSNKALTZY---EDVSPAPDNQIGMNTAGVWNTVPPSQ 794
 ||| : : : : : ||| : : : : :
 Db 783 YELKV---QAVNQAGGPEQVITIGSGEDYQVQSPLEDDITFNSTVLVWRKPDVLA 838
 ||| : : : : : ||| : : : : :
 Qy 795 HHNGNLGYKI-----EVSAGNMKVLNMTLWATTSVLLNNLTGAVTSVRLNS 845
 ||| : : : : : ||| : : : : :
 Db 839 QVKGHLGASVITWYNNKSGRKHSHIK--SHIVVPAMTTSAILSGRPFYSVHYEVA 896
 ||| : : : : : ||| : : : : :
 Qy 846 PTKAGDGP---YSKPSISLMOPH-----HVHPHRAHPSGTHDG---881
 ||| : : : : : ||| : : : : :
 Db 897 FNGKLGAPASWTFSTPEVY---PGHPEALHLCQSDTSLHLLWQPGSLH-NOLYGLLL 952
 ||| : : : : : ||| : : : : :
 Qy 882 -RH--EQQ-----DLTYHNNGNIPGDNIPHTHKKTITDYLSPMLWLCIVL 926
 ||| : : : : : ||| : : : : :
 Db 953 SYHPVGESEKQFNLSDPELTHNLNLP-DLYRHFQLOAQVQGGGHEAIVREGGT 1011
 ||| : : : : : ||| : : : : :

Qy 927 LVLV-----ISA-----AISMYFK-----RHQMTKELGHSVSDNEITALIN 967
 ||| : : : : : ||| : : : : :
 Db 1012 WALGAGPDGMSIAGENTSYSVSVPRKQCNRFRHLKALPBKGVSPDHPQVQVVS 1071
 ||| : : : : : ||| : : : : :
 Qy 968 SKBSLWIDHHRGRW-TADTDKDGSLSESKILLSVNSQSNYSNDSGGDTYAEVDT 1021
 ||| : : : : : ||| : : : : :
 Db 1072 YNQSSTQ---WNLQPDTKYIEHLKELVLLHLDVKTN-----GTGPRVST 1116
 ||| : : : : : ||| : : : : :

RESULT 9
 DCC_MOUSE
 ID DCC_MOUSE STANDARD; PRT: 1447 AA.
 AC P70211;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
 GN DCC.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=BRAIN;
 RX MEDLINE=96112625; PubMed=6570174;
 RA Cooper H.N., Ames P., Britto J., Gad J., Wilks A.F.;
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer
 RT gene (mDCC) and its expression in the developing mouse embryo."
 RL Oncogene 11:2243-2254(1995).
 RN [2]
 RP REVISIONS.
 RC STRAIN=BALB/C; TISSUE=BRAIN;
 RA Cooper H.N.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
 CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
 CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
 CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
 CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION, HIGHEST LEVELS
 CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
 CC AND REMAIN AT THIS LEVEL IN THE ADULT.
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X85788; CAA59786.1; .
 DR HSP: P56276; 17LK.
 DR MG: M0194689; DCC.
 DR INTERPRO: IP0001777; .
 DR INTERPRO: IP0003006; .
 DR PFAM: PF00041; fn3; 6.
 DR PFAM: PF00047; Ig; 4.
 DR PRINTS: PR00014; FNTYPEIII.
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KW Anti-oncogene; Alternative initiation; Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, LONG
 FT ISOFORM.
 FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, SHORT
 FT ISOFORM.
 FT INIT_MET 85 85 FOR SHORT ISOFORM.

RA Oliner J.D., Kinsler K.W., Vogelstein B. ;
 RT "Scrambled exons." ;
 RL Cell 64:607-613(1991).
 RN [4]
 RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
 RX MEDLINE=94245241; PubMed=8188295;
 RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Pearson E.R.,
 RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B. ;
 RT "The DCC gene: structural analysis and mutations in colorectal
 RT carcinomas." ;
 RL Genomics 19:525-531(1994).
 RN [5]
 RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
 RX MEDLINE=94243823; PubMed=8187090;
 RA Miyake S., Nagai T., Yoshino K., Ota M., Endo M., Yuasa Y. ;
 RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
 RT human esophageal squamous cell carcinomas and their relation to
 RT metastasis." ;
 RL Cancer Res. 54:3007-3010(1994).
 CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
 CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
 CC -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
 CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
 CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
 CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
 CC METASTASIS OF ESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
 CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: X76132; CAA53735.1 ;
 DR EMBL: M12292; AAA35751.1 ;
 DR EMBL: M12286; AAA52174.1 ;
 DR EMBL: M12288; AAA52175.1 ; ALT_SEQ.
 DR EMBL: M12290; AAA52176.1 ;
 DR EMBL: M63696; AAA52177.1 ;
 DR EMBL: M63700; AAA52178.1 ;
 DR EMBL: M63702; AAA52179.1 ;
 DR EMBL: M63718; AAA52180.1 ;
 DR EMBL: M63698; AAA52181.1 ;
 DR PIR: A54100; A54100.
 DR PIR: A40098; A40098.
 DR PIR: A38442; A38442.
 DR HSP: P56276; I7LK.
 DR MIM: 120470 ;
 DR INTERPRO: IP0001777 ;
 DR INTERPRO: IP0003006 ;
 DR PFAM: PF00041; fn3; 6.
 DR PFAM: PF00047; ig; 4.
 DR PRINTS: PR00014; FNTYPEIII.
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KW Anti-oncogene; Disease mutation; Polymorphism.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC.
 FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1098 1122 POTENTIAL.
 FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 54 124 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 154 219 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 254 317 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 426 522 FIBRONECTIN TYPE-III.
 FT DOMAIN 525 618 FIBRONECTIN TYPE-III.
 FT DOMAIN 619 716 FIBRONECTIN TYPE-III.

FT DOMAIN 722 816 FIBRONECTIN TYPE-III.
 FT DOMAIN 840 940 FIBRONECTIN TYPE-III.
 FT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
 FT DISULFID 61 117 BY SIMILARITY.
 FT DISULFID 161 212 BY SIMILARITY.
 FT DISULFID 261 310 BY SIMILARITY.
 FT DISULFID 352 400 BY SIMILARITY.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 168 168 M -> T (IN ESOPHAGEAL CARCINOMA).
 FT VARIANT 201 201 R -> G.
 FT VARIANT 1375 1375 P -> H (IN A COLORECTAL CARCINOMA).
 FT VARIANT 1375 1375 P -> H (IN A COLORECTAL CARCINOMA).
 FT CONFLICT 138 138 MISSING (IN REF. 3).
 FT CONFLICT 233 329 MISSING (IN REF. 3).
 FT CONFLICT 421 421 MISSING (IN REF. 3).
 SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;

Query Match 8.6%; Score 637.5; DB 1; Length 1447;
 Best Local Similarity 22.94; Pred. No. 1.4e-24;
 Matches 271; Conservative 150; Mismatches 444; Indels 317; Gaps 43;

Qy 57 RIIEHPDLVKKNEPATLNCVKG-KPEPTIDWFGK--EPVSTNEKSHVQPKDGL 113
 Db 41 RFLSEDPVAVMKGWVLLDCAESDRGVVVKMKDGLHGLMDGRKQ--QLSGSL 97
 Qy 114 FFRYTKQKQKQ-DGGETVCVAV-NRVQANVSHASQLA-VLADDPVPEKDTVRAGE 170
 Db 98 LIQNLASHKPKQDGLYQCEASDGSSTISRTAVVAVPLR--FLSQSTVSTAFMG 155
 Qy 171 TALLECGPKPKIPEPTLIWIKDGVPLDLKAMSGFASNRVIRVGGNLSINVEPIDGN 230
 Db 156 TVLLKCV-EVIGSEPMPTIHWKMKQDLTPIP-----GDSRVVFLSGDLQISLQ 209
 Qy 231 YKCIATNGLTRESSTAKLVQKPK-----FYKPKQKQVLLGQTATFHCSDGPP 284
 Db 210 YKCSANPASPSTGNDAEVRILSDGLRLQVLPQSRVVALBGKDAVLECCVSTP 269
 Qy 285 KVMKKKEGNIPV-SRARIHLDEKSEISNITPDGETGYCAENNVQISARSLVHA 343
 Db 270 SPTVLGRSEVILQSRKYSLGSGNLLSMTVDSDSGMTCTVTKNENISASAEILV 329
 Qy 344 PNPPTKPSNKGVLGVQLPCMASGPPPSVFWTEKGVSTLMPNSHGKQVAVDGT 403
 Db 330 PPNFLHNSM-----LYATESM----- 346
 Qy 404 LQITDVRQDGETTYCYSAFVSVDSTVRFVQLVSGVSDRPPPIIQQPANTL 463
 Db 347 ---DIEFE----- 351
 Qy 464 LPRCATGSPRIKWHGDAVQAGNRYSIQSSLRVLDQLSDSTTCTASGERGET 523
 Db 352 --CTVSGVPVTVNMKNKVDVPSDYFQIVGGSNRLTGLVSGDEGPTQCVARNEGNA 409
 Qy 524 SWAATLIVKPGSTSLHRAEDPTSPATPOTPKVLNVSRTISLWAKSQXKPGVPII 583
 Db 410 QTSAGLIVKTPAIS-----SSVLPAPRDVPLVSVSRVFLSWRPAPAE--AGNIQ 460
 Qy 584 GTIVTFSPD-----LQTGVAAHVHVDQTVISGLTPTGSTVFLVRAENTQISVP 636
 Db 461 TPTV-FPSREDNREALAT-----TOPGSLITVGNLKPAMYTRVAVNENG--P 510
 Qy 637 SGLSNVITIEDADPASANDLSARTLLVKSGLVSDASLNASVLRNWLHVSDEK 696
 Db 511 GESSQPIK-----VATPELQVPGVENVLQAVSTPTSLITWEPYANGPP 557

Qy 697 TVEGLRIKDYASVPSAQYHSITVMDASAEFVVGKLYKATYKTEFFLTPFFETBQSPQN 756
 ||| : : : | : : : | : : : | : : : | : : :
 Db 558 -VGGVFLCTEYSGKEDQ-----LEVGLSKYLEGKAKTETSLRFLATNRKGVGVSVD 611
 Qy 757 SKTALTYEDVSPAPDNIQIGMYNQTAGVYMTPPSQHNGNLGYKI---EVSAGWM 813
 | : | | | | | : : : | : | | | | : : | : | : : :
 Db 612 DITVTVLSGVSPAPQNVSLVYNSRSIKVSLPFPSTGQNGTGVKIRHKRTTRGEM 671
 Qy 814 KVLNLTNATTSVLLNLTGAVYVRLNSPKTAGDQPSKPSILPMDPTHVHPFPA 873
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 672 -TLEPNMLVLTGLKXGQYSPQVSAMTVNGTP----- 706
 Qy 874 HPSGTEHGRHGGDLTYNNGNIP--PGD--INPTTHKATTDYSGPWMLVLCVLVL 929
 : : : : | : : : | : : : | : : : | : : :
 Db 707 -PSWVTAETPNDL--DESGVDPQPSLHVPOTN-----CIIM-- 743
 Qy 930 VISAAISMVYFKRKHQMKELGLSVSNDEITALMNS--KESLWIDHRGHWATDIX 987
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 744 -----SWTPPL-NPKIVVGTGIVGVGSPVTVVSDQKRYTSIERLE 787
 Qy 988 DSGLSKSLSHVNSGQSNVNSDGGD-YAEVDRLNTFTNCRKSPNPTPYATMII 1046
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 788 SS-----SEIVISLAKFNAGSGVPLYSATRSIT--DPTDPVY-- 826
 Qy 1047 GTSSTCTCTKTSISADKDSGTHSPYDAFQAPV-----PV-VKSNLYQVPEIN 1099
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 827 -----YPLDDPTSPVDSLTPMLPPGVQVVALHDVAVWS 863
 Qy 1100 WSEFLPPPPPPPPSYTYGAQSGPESSKSKSAGSGISTQSLMASTHSSSSGGFSA 1159
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 864 WADSWPKMWTSEVRLYVWRWTSFSAAYKX-----EDTSLSTYA 907
 Qy 1160 WGVSP-----QVAVACPPENVYNPLSAVAGQTQNRQIIPIN 1197
 | : | : | : | : | : | : | : | : | : | : | : | :
 Db 908 TGLAFVWVWFESVYVW-TKNRSSSTWSTAHAT--TYEAPTS 946

RESULT 11
 CAML_HUMAN
 ID CAML_HUMAN STANDARD; PRT; 1257 AA.
 AC p32004;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE LI PRECURSOR (N-CAM LI).
 GN LICAM OR CAMLI OR NICS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mollusca; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92031698; PubMed=1923117;
 RA Kobayashi M., Miura M., Asou R., Oyemura K.;
 RT "Molecular cloning of cell adhesion molecule LI from human nervous
 RT tissue: a comparison of the primary sequences of LI molecules of
 RT different origin.";
 RL Biochim. Biophys. Acta 1090:238-240 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rosenthal A., Coutelle O., Drescher B.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92329299; PubMed=1627459;
 RA Reid R.A., Hemperly J.J.;
 RT "Variants of human LI cell adhesion molecule arise through alternate
 RT splicing of RNA.";
 RL J. Mol. Neurosci. 3:127-135 (1992).
 RN [4]
 RP SEQUENCE OF 353-1176 FROM N.A.
 RX MEDLINE=92020233; PubMed=1923824;
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
 RT "PCR walking from microdissection clone M54 identifies three exons

RT from the human gene for the neural cell adhesion molecule LI
 RT (CAM-LI).";
 RL Nucleic Acids Res. 19:5395-5401 (1991).
 RN [5]
 RP SEQUENCE OF 332-371 FROM N.A.
 RX MEDLINE=90353957; PubMed=2387585;
 RA Djaheli M., Matti M.-G., Nguyen C., Roux D., Demengeot J.,
 RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
 RT "The gene encoding LI, a neural adhesion molecule of the
 RT immunoglobulin family, is located on the X chromosome in mouse and
 RT man.";
 RL Genomics 7:587-593 (1990).
 RN [6]
 RP SEQUENCE OF 1030-1257 FROM N.A.
 RX MEDLINE=91132183; PubMed=1993895;
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
 RA Stallcup W.B.;
 RT "Isolation and sequence of partial cDNA clones of human LI: homology
 RT of human and rodent LI in the cytoplasmic region.";
 RL J. Neurochem. 56:797-804 (1991).
 RN [7]
 RP SEQUENCE OF 20-36.
 RX MEDLINE=88298876; PubMed=3136168;
 RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
 RA Rathjen F.G.;
 RT "A human brain glycoprotein related to the mouse cell adhesion
 RT molecule LI.";
 RL J. Biol. Chem. 263:11943-11947 (1988).
 RN [8]
 RP VARIANT HSAS TYR-264.
 RX MEDLINE=94004956; PubMed=8401576;
 RA Jouet M., Rosenthal A., Macfarlane J., Kenrick S., Donnai D.;
 RT "A single missense mutation confirms the LI defect in X-linked hydrocephalus
 RT (HAS).";
 RL Nat. Genet. 4:331-331 (1993).
 RN [9]
 RP VARIANT HSAS/MASA LEU-1194.
 RX MEDLINE=95187172; PubMed=7881431;
 RA Fransen E., Schrandt-Stumpel C., Vits L., Coucke P., van Camp G.,
 RA Willems P.J.;
 RT "X-linked hydrocephalus and MASA syndrome present in one family are
 RT due to a single missense mutation in exon 28 of the LICAM gene.";
 RL Hum. Mol. Genet. 3:2255-2256 (1994).
 RN [10]
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
 RX MEDLINE=95004608; PubMed=7920659;
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
 RA Peterson J., Mettenberg A., Ionescu V., Temple K., Kenrick S.;
 RT "X-linked spastic paraplegia (SPGL), MASA syndrome and X-linked
 RT hydrocephalus result from mutations in the LI gene.";
 RL Nat. Genet. 7:402-407 (1994).
 RN [11]
 RP VARIANTS MASA GLN-210 AND ASN-598.
 RX MEDLINE=95004609; PubMed=7920660;
 RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
 RA Reymers E., Korn B., Poustka A., Wilson G., Schrandt-Stumpel C.,
 RA Winter R.M., Schwartz C., Willems P.J.;
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene
 RT LICAM.";
 RL Nat. Genet. 7:408-413 (1994).
 RN [12]
 RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
 RX MEDLINE=95282776; PubMed=7762552;
 RA Jouet M., Mondie A., Peterson J., McKeown C., Fryer A., Carpenter N.,
 RA Holmberg E., Wadellius C., Kenrick S.;
 RT "New domains of neural cell adhesion molecule LI implicated in
 RT X-linked hydrocephalus and MASA syndrome.";
 RL Am. J. Hum. Genet. 56:1304-1314 (1995).
 RN [13]
 RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
 RX MEDLINE=96153146; PubMed=8556302;
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,

RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
 RT due to mutations in one single gene. Li.
 RL Eur. J. Hum. Genet. 3:273-284 (1995).
 RN [14]
 RP ERRATUM.
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RL Eur. J. Hum. Genet. 4:126-126 (1996).
 RN [15]
 RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
 RX MEDLINE-960571; PubMed-7562969;
 RA Ruiz J.C., Cuppens H., Legius E., Frys J.-P., Glover T., Marynen P.,
 RA Cassiman J.-J.;
 RT "Mutations in LI-CAM in two families with X linked complicated
 RT spastic paraplegia, MASA syndrome, and HSAS";
 RL J. Med. Genet. 32:549-552 (1995).
 RN [16]
 RP VARIANTS HSAS CYS-194 AND LEU-240.
 RX MEDLINE-97083370; PubMed-8929944;
 RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
 RA Engel W., Schwinger E., Gal A.;
 RT "Five novel mutations in the LICAM gene in families with X linked
 RT hydrocephalus";
 RL J. Med. Genet. 33:103-106 (1996).
 RN [17]
 RP VARIANTS HSAS Q-184; V-439-T-443 DEL; C-784 AND L-936-L-948 DEL.
 RX MEDLINE-97338664; PubMed-9195224;
 RA Macfarlane J.R., Du J.-S., Peygs M.E., Ramsden S., Donnai D.,
 RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
 RA Moncla A., Lunt P., Hodgson S., Jovet M., Kenwick S.;
 RT "Nine novel LI CAM mutations in families with X-linked
 RT hydrocephalus";
 RL Hum. Mutat. 9:512-518 (1997).
 RN [18]
 RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
 RX MEDLINE-98107021; PubMed-9521424;
 RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
 RT "Multiple exon screening using restriction endonuclease
 RT fingerprinting (RFP): detection of six novel mutations in the LI cell
 RT adhesion molecule (LICAM) gene";
 RL Hum. Mutat. 11:222-230 (1998).
 RN [19]
 RP VARIANT CRASH PRO-632.
 RX MEDLINE-98112489; PubMed-9452110;
 RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
 RA Willems P.J.;
 RT "Evidence for somatic and germline mosaicism in CRASH syndrome";
 RL Hum. Mutat. Suppl. 1:S284-S287 (1998).
 RN [20]
 RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
 RX MEDLINE-98415726; PubMed-9744477;
 RA Saugier-Veber P., Martin C., le Mear N., Lyonnet S., Munnich A.,
 RA David A., Benocq A., Heron D., Jouveaux P., Odent S., Manouvrier S.,
 RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
 RT "Identification of novel LICAM mutations using fluorescence-assisted
 RT mismatch analysis";
 RL Hum. Mutat. 12:259-266 (1998).
 CC #1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO ALEXIN ON NEURONS.
 CC #1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC #1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING.
 CC #1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF THREE X-LINKED
 CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
 CC SYLVUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
 CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
 CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
 CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
 CC Milder presentation and a longer life expectancy. 3: SPASTIC
 CC paraplegia type 1 (SPG1). COLLECTIVELY THESE SYNDROMES ARE ALSO
 CC KNOWN AS CRASH SYNDROME, AN ACRONYM WHICH STANDS FOR CORPUS
 CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,

CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
 CC #1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF HIRSCHSPRUNG DISEASE
 CC (HSCR).
 CC #1- SIMILARITY: CONTAINS 6 IMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC #1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC #1- DATABASE: NAME-LICAM; NOTE-LICAM mutation Web Page;
 CC WWW:"http://hgins.uia.ac.be/dnalab/l1".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X59847; CAA42508.1; .
 DR EMBL: Z29373; CAA82564.1; .
 DR EMBL: M74387; CAA59476.1; .
 DR EMBL: X58775; CAA41576.1; .
 CC
 CC Query Match 8.5%; Score 633.5; DB 1; Length 1257;
 CC Best Local Similarity 23.7%; Pred. No. 1.8e-24;
 CC Matches 243; Conservative 142; Mismatches 395; Indels 245; Gaps 36;
 CC
 QY 35 WLLLVLLVAVGSLPVAHQVQSPRIE-----HPTDLVYKKNPAPLAKVKEGPEPTT 87
 Db 9 WLLLVLLVAVGSLPVAHQVQSPRIE-----HPTDLVYKKNPAPLAKVKEGPEPTT 87
 QY 88 ENKDGEPVSTNEKSHRVQFDGALFPYTMGRK-----EQDGEYVCAVKNRQVQV 143
 Db 68 ENKDGEPVSTNEKSHRVQFDGALFPYTMGRK-----EQDGEYVCAVKNRQVQV 143
 QY 144 RHASLIQVLDNRQVFEVPEPT-----RVAKETALLECGPPKGIPEPLTWIMQVPLD 199
 Db 126 H-----SILMAEGAPKWPATVTKVPEVGEBSVCLVNPCCSAPLRTIMMS----- 174
 QY 200 KAMSPGASSRVLVVGNGLLISN-----EPI----- 226
 Db 175 KILHILKQDERTVMQNGNLIYFANLISDWSYDIAHAFPTRTTIQKPIDLRVAKNS 234
 QY 227 ----- 226
 Db 235 WIDRKRPLLFPTNSSSHLVALGQPLVLECIAGFPPTIKWLKPSGMPADRVPTQNN 294
 QY 227 -----DBGNKTCIAQNLVGTRESSYALIVQVPIFMKPKDQVLMIGQTAIFR 275
 Db 295 KTLQGLKVEGSDGQYRCIAENSLGASRAHY-VTVYVAAPVWLKQPSHLLPGGTARLD 353
 QY 276 CSVGQDPFPPKVLWKEGNIIPVSRARILDEK-----SLEISNITPDGTVCAHNN 329
 Db 354 CQVQSGPQFPTWR--INGIPE--ELAKQYRIQRGALLISQVQSDTHVTOCAHNN 409
 QY 330 VQGISARSL-IIVHAPNPTKPSKNKVLGNG-VQPLCMAAGNPPSPVFTKEGVTLM 387
 Db 410 HGLLIANVITVQVQKALITADQNTYMAVQGSTALILCKARFAGVPSQWLEDDQTVL 469
 QY 388 FPNSSHGRIYVADQTLQITDVRQEDGYGVSAFVSQSVSTVRFVLSVSDERPPPI 447
 Db 470 ----QDERFFPTANGILRLQDQNTGYRCIAAQQDQNTVIMANKVDAQT-----I 520
 QY 448 QIGPANTQLFAGSVATLPRATGNPS--PRIKHVHGHQVQA--GNRTSIQGSGLRVD 502
 Db 521 TQGPSTIKGQSRVPTTQCAQSPDLSIPTITWGGDGLQELGDSQK--FIEDGLVTH 579
 QY 503 DLQLSQSGTITCTASGERGT-SWAATLTVKPGSTLSHRAADPTTPAGPTPKVLN-V 560
 Db 580 SLQTSQNTCTVASTELDVESRQLVVGSGP-----PPVRLVLSDEHLL 626
 QY 561 SRTSISLAWAKSQKQAGVPIGTIVTGYSPDL-QTGVVIAHRVGDQTVISILPGT 619
 Db 627 TQSQVRVMSQAEHNN--APIEKVDFEDKEMAKPEKSVSLGKVPNGTSTLKLSPV 683


```

CC      GPI-ANCHOR.
CC      *- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC      *- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC      (-----)
CC      THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION-
CC      THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC      ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC      (-----)
DR      ENGL: X63101; CAA44815.1; -.
DR      PIR: S22128; S22128.
DR      PIR: S22383; S22383.
DR      HSSP: P56726; 1TLA.
DR      INTERPRO: IPR001777; -.
DR      INTERPRO: IPR003006; -.
DR      PFAM: PF000041; fn3; 3.
DR      PFAM: PF00047; I9; 6.
KW      Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW      Cell adhesion; Repeat.
FT      SIGNAL      1      23      OR 25 (POTENTIAL).
FT      CHAIN       24: 1036      AKONIN-1.
FT      PROPEP      7: 1036      REMOVED IN MATURE FORM.
FT      DOMAIN      49      113      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      143     211      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      249     308      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      336     397      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      428     490      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      518     589      IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      599     608      HINGE (POTENTIAL).
FT      DOMAIN      601     607      GLY/PRO-RICH.
FT      DOMAIN      608     709      FIBRONECTIN TYPE-III.
FT      DOMAIN      710     811      FIBRONECTIN TYPE-III.
FT      DOMAIN      812     912      FIBRONECTIN TYPE-III.
FT      DOMAIN      913     1009     FIBRONECTIN TYPE-III.
FT      MCD_RES     924      924      BLOCKED.
FT      CARBOHYD    71      71      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    199      199      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    456     456      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    472     472      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    493     493      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    520     520      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    770     770      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    900     900      N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD    914     914      N-LINKED (GLCNAC. .) (POTENTIAL).
SO      SEQUENCE    1036 AA; 113301 MW; 08B80143B8779794 CRC54;

```

LD AXOL_CHICK STANDARD; PRT: 1036 AA.
AC P26855;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR.
CS Gallus gallus (Chicken).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=BRAIN;
RX MEDLINE=92174898; PubMed=1311675;
RA Zuelli R.A., Rader C., Schroeder A., Kalousek M.B.,
RA von Bohlen Und Halbach P., Osterwalder T., Ioan C., Stoeckli E.T.,
RA Affolter H.-U., Fritz A., Hafen E., Sondergerger P.;
RT "The axonally secreted cell adhesion molecule, axonin-1. Primary
RT structure, immunoglobulin-like and fibronectin-type-III-like domains
RT and glycosyl-phosphatidylinositol anchorage."
RL Eur. J. Biochem. 204:453-463(1992).
CC -1- FUNCTION: AXON-ASSOCIATED CELL ADHESION MOLECULE (AXCAM) WHICH
CC PROMOTES NEURITE OUTGROWTH BY INTERACTION WITH THE AXCAM L1 (G4)
CC OF NEURITIC MEMBRANE.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A

```

Query Match          8.4%; Score 626.5; DB 1; Length 1036;
Best Local Similarity 24.2%; Pred. No. 3.2e-24;
Matches 241; Conservative 130; Mismatches 390; Indels 233; Gaps 35;

QY 71 EPATLNCVKVEKPEPTIEWFKDGPESVTEKKSHRVQDKGALFFYRTMQGK ----- K 123
Db 50 EKVLITLCARANPATRN-----KNNGTELKMGDPSRSLVLADGLVSNIPYK 97

QY 124 BQDQGETCYCAKVRVQAVSRHASQIAVLRDQFVEFKD-TRVAKGETALLBGGPKGI 182
Db 98 AKAGDSQTCYATNARQTVSRASLRFGFLQ- EFSASRDEPVKITBGGWMTCSPPHY 156

QY 183 PEPTLIWIKDQVDPGLDLKAMSPGASSNRVIND- GGNLLINSNPEIDBGNKYCTAQLNG 240
Db 157 PALSTWNLHLEF-----NPIADGRGVRSGTGNLIARTASDGLNGSCFATSHD 209

QY 241 ---THE-SSVAKVIL-----QWPKF-MKEPDQVMQLQATPFCSGVGGSDPPVVLAK 289
Db 210 FIKVSVSKFSQSLSLAEDARDAPSAIKAPADYATLTQWTLCLCFARNPQVFNKR 269

QY 290 KEGNPIVSRARILHDEKSLSPINPTDDEPTCEAHNNGVQVSLALSHAPVNPET 349
Db 270 KLDGS---QSKSLSSPELLHVDQVEDPTEGTECEANTKGRQCTYGGRIILHAPQDMLD 326

```



```

QY 771 PDNIQIYQKTDAGWYVPPSPSHHNGILYGIK-----EVSAGNMYKLAN-MTLAA 823
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 825 PGNVGVYVNSTIAKRVWDPVPLASVGHGLQYKYYTKVQSLRSRSGRHEKKILFTFG 884
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 824 TTTSVLLNLLITGAVISVRLNPSKTACGDYSPKISLFDMPDTHVHPVR----- 872
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 885 NKTGCMPLGPEPSTSLKRVNVVNGKGGCPA-PDKVFETPGCVSPSPSFLKITNPLDS 943
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 873 -----AHSPG 877
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 944 IITLWMCSTPTWPG 956
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 15
CONT_HUMAN
ID CONT_HUMAN STANDARD; PRT; 1018 AA.
AC Q12861; Q12861; Q14030
DT 01-NOV-1997 (Rel. 35; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE CONTACTIN PRECURSOR (GLYCOPROTEIN GPI35).
GN CNTN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RW [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=BRAIN;
RA MEDLINE=95048335; PubMed=7959734;
RE Berglund E.O., Ranscht B.;
RT "Molecular cloning and in situ localization of the human contactin
RT gene (CNTN1) on chromosome 12q11-q12."
RL Genomics 21:571-582(1994).
RW [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC MEDLINE=94217459; PubMed=8164510;
RA Reid R.A., Hemperly J.J.;
RT "Identification and characterization of the human cell adhesion
RT molecule 'contactin'."
RL Brain Res. Mol. Brain Res. 21:1-8(1994).
CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
CC SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2 ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a colla-
CC boration between the Swiss Institute of Bioinformatics and the EMBL out-
CC side the European Bioinformatics Institute. There are no restrictions
CC on use by non-profit institutions as long as its content is in
CC modified and this statement is not removed. Usage by and for com-
CC mercial entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 22, 2001, 12:50:23 ; Search time 559.88 Seconds
(without alignments)
292.036 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPMPEHNAHARSTSTNN.....SCLTAEAGEAPAPROMTANN 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMGL_15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organella.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertibrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	7427	100.0	1395 5	O44924
2	7407	99.7	1395 5	Q9W213
3	1620	21.8	859 5	Q9V226
4	1609.5	21.7	1273 5	O44928
5	1607.5	21.6	1612 11	O89026
6	1592	21.4	1651 4	Q9Y6N7
7	1585	21.3	1651 11	O55005
8	1499	20.2	1060 11	Q9QZ13
9	1430	19.3	823 5	Q9VQ10
10	1395.5	18.8	1344 11	Q92274
11	826.5	11.1	423 5	P91572
12	790	10.6	874 5	O01632
13	759.5	10.2	1377 11	P97603
14	731.5	9.8	1493 11	P97798
15	712.5	9.6	1461 4	O00340
16	710.5	9.6	1461 4	Q92859
17	686	9.2	1277 13	Q98902
18	677	9.1	1443 13	Q90610
19	667.5	9.0	1822 4	Q9UL77

20	663.5	8.9	1026 4	Q94780	Q94780 homo sapien
21	659.5	8.9	1100 4	Q94779	Q94779 homo sapien
22	652	8.8	2016 5	Q9V479	Q9V479 drosophila
23	651	8.8	1026 11	Q62845	Q62845 rattus norv
24	651	8.8	2221 5	Q9U1M1	Q9U1M1 drosophila
25	649	8.7	2016 5	Q9VBA1	Q9VBA1 drosophila
26	647	8.7	1028 11	Q62682	Q62682 rattus norv
27	645.5	8.7	1445 11	Q63155	Q63155 rattus norv
28	645	8.7	1028 11	Q07409	Q07409 mus musculu
29	644	8.7	1272 13	Q90924	Q90924 gallus gall
30	643	8.7	1217 11	P97685	P97685 rattus norv
31	636	8.6	2222 5	Q97394	Q97394 drosophila
32	635	8.5	1099 11	P97527	P97527 rattus norv
33	632	8.5	1264 5	P91767	P91767 manduca sex
34	631	8.5	1369 13	Q42414	Q42414 gallus gall
35	629	8.5	1259 11	Q9QY38	Q9QY38 mus musculu
36	628	8.5	1028 11	P97528	P97528 rattus norv
37	626.5	8.4	1248 6	Q9XT41	Q9XT41 cercopithec
38	624	8.4	1028 11	Q9JMB8	Q9JMB8 mus musculu
39	622.5	8.4	1427 13	Q91562	Q91562 xenopus lae
40	619.5	8.3	1018 6	Q28106	Q28106 bos taurus
41	616.5	8.3	1302 5	O61542	O61542 drosophila
42	616	8.3	1166 11	Q9QVNA	Q9QVNA rattus sp.
43	612.5	8.2	1151 11	Q9QVNS	Q9QVNS rattus sp.
44	612.5	8.2	1215 11	P97686	P97686 rattus norv
45	611	8.2	1239 5	Q9V3X0	Q9V3X0 drosophila

ALIGNMENTS

RESULT 1
O44924
ID O44924 PRELIMINARY; PRT; 1395 AA.
AC O44924;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Bpilydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors."
EL Cell 92:205-215(1998).
DR ENBL: AF040989; AAC38849.1; .
DR HSP: P56276; 1TLK.
DR FLYBASE: FBgn0035631; robo.
DR INTERPRO: IPRO01777; .
DR INTERPRO: IPRO03006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; lg; 5.
DR PRINTS: PR00014; FWTPEI1.
SQ SEQUENCE 1395 AA; 151778 MW; B820E234A5218983 CRC64;

Query Match 100.0%; Score 7427; DB 5; Length 1395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHPMPEHNAHARSTSTNNPFRSSRSRWLLPAMLLLVASMLPAVQGLQSPRIIE 60
|||||
Db 1 MHPMPEHNAHARSTSTNNPFRSSRSRWLLPAMLLLVASMLPAVQGLQSPRIIE 60
|||||
Qy 61 HPTDLVKKNCNPNALCKVKGKPEPTINFEKDGEPVSTNEKSHRVQFKDGLAFYTRMQ 120

Db 61 HPTDVLKNEPATLCKVGEKPEPTIEWPKDGPVSTKKAHSHPQDKGALFFRTMQ 120
Qy 121 GKKQDGGRTVCAKNRVQVASHASQIAVLARDRFRVPEPTVAKGATLALCGPPK 180
Db 121 GKKQDGGRTVCAKNRVQVASHASQIAVLARDRFRVPEPTVAKGATLALCGPPK 180
Qy 181 GIPETPLTKIKGVPLDILKAMSGASSRVIRVGGNLLSNWPEIDENGYCIAQLVNG 240
Db 181 GIPETPLTKIKGVPLDILKAMSGASSRVIRVGGNLLSNWPEIDENGYCIAQLVNG 240
Qy 241 TRESSYAKLIVQVKPYFKPKDQVLMVIGQATPCFSGVDGPPVLMKKKEGNIPIVSA 300
Db 241 TRESSYAKLIVQVKPYFKPKDQVLMVIGQATPCFSGVDGPPVLMKKKEGNIPIVSA 300
Qy 301 RILHDEKSLISNITPTDGTTCVCAHNVQGISARASLIVHAPPNFKRPSMKVGLNG 360
Db 301 RILHDEKSLISNITPTDGTTCVCAHNVQGISARASLIVHAPPNFKRPSMKVGLNG 360
Qy 361 VQPLQKASGNPPSPVFWIKGVSTLMFNSHSGROVYADGQITITDVRDEGTVC 420
Db 361 VQPLQKASGNPPSPVFWIKGVSTLMFNSHSGROVYADGQITITDVRDEGTVC 420
Qy 421 AFSWDSSTVRFVLQVSSYDERPPPIIGQIPANQITPKGSVATLPCRATGNPSPIKKNH 480
Db 421 AFSWDSSTVRFVLQVSSYDERPPPIIGQIPANQITPKGSVATLPCRATGNPSPIKKNH 480
Qy 481 DGHVAGWAGNYSIIQSSSLAVDDQLSDSGYITCAGSERGTSNAAVLVEKPSSTLSH 540
Db 481 DGHVAGWAGNYSIIQSSSLAVDDQLSDSGYITCAGSERGTSNAAVLVEKPSSTLSH 540
Qy 541 RAADPTSPAPPTGPKVLNVKSTISLMAKSOEKPGAVGPIIGTVETVSPDQGWIV 600
Db 541 RAADPTSPAPPTGPKVLNVKSTISLMAKSOEKPGAVGPIIGTVETVSPDQGWIV 600
Qy 601 AAHVQDQDTQISGLTPTSTVFLVRAETQGISVPSGLSNVKKIIEADPDAASNDLSA 660
Db 601 AAHVQDQDTQISGLTPTSTVFLVRAETQGISVPSGLSNVKKIIEADPDAASNDLSA 660
Qy 721 ARTLLGKSVELIDASINASAVRLKMLVSADEKVEGLRIHKYDASVPSQAHSITV 720
Db 721 ARTLLGKSVELIDASINASAVRLKMLVSADEKVEGLRIHKYDASVPSQAHSITV 720
Qy 721 MDAESAFTVGNLAKYTKTEFFLPFFETIGQPSNKATLITVDPVAPPDNIQGMN 780
Db 721 MDAESAFTVGNLAKYTKTEFFLPFFETIGQPSNKATLITVDPVAPPDNIQGMN 780
Qy 781 QTAGWVWPPSPQHNGNLGTYKIEVAGMNTVLAJNLNATTVSLVNLNLTGAVS 840
Db 781 QTAGWVWPPSPQHNGNLGTYKIEVAGMNTVLAJNLNATTVSLVNLNLTGAVS 840
Qy 841 VNLSPFTKAGDQSKPISLMDPTHHVHPRAHPSGTGHDGRHQDQLTYHNNGNIPPD 900
Db 841 VNLSPFTKAGDQSKPISLMDPTHHVHPRAHPSGTGHDGRHQDQLTYHNNGNIPPD 900
Qy 901 INPTTHAKTITDLSGPMWMLVLCVLLVLVISAISMYKPKRKHMTKELGHSVSDNE 960
Db 901 INPTTHAKTITDLSGPMWMLVLCVLLVLVISAISMYKPKRKHMTKELGHSVSDNE 960
Qy 961 ITALNINSGKSLVDHGRWRTADTKDGLSEKSLSHVNSGSGNTNSGGDTTARVD 1020
Db 961 ITALNINSGKSLVDHGRWRTADTKDGLSEKSLSHVNSGSGNTNSGGDTTARVD 1020
Qy 1021 TNLNLTPTNCRGSPDNPTIATMTIGTSSECTKTITSIAKDQSGHSPISDAFAGQV 1080
Db 1021 TNLNLTPTNCRGSPDNPTIATMTIGTSSECTKTITSIAKDQSGHSPISDAFAGQV 1080
Qy 1081 PAVPVKSNVLTQVPEVFNSEFLPPHPPPPSSSTGYAQGPSSEKSSKSGAGSIST 1140
Db 1081 PAVPVKSNVLTQVPEVFNSEFLPPHPPPPSSSTGYAQGPSSEKSSKSGAGSIST 1140
Qy 1141 NQSLINASIHSSSSGGFSAWGSQYAVACPPENYSNPLSAVAGGTQNTQITPTNQRP 1200

Db 1141 NQSLINASIHSSSSGGFSAWGSQYAVACPPENYSNPLSAVAGGTQNTQITPTNQRP 1200
Qy 1201 PQLPATFATTGGGAVPNHLPFATROHAASEYQAGLNARCAQSRACNSDALATPSM 1260
Db 1201 PQLPATFATTGGGAVPNHLPFATROHAASEYQAGLNARCAQSRACNSDALATPSM 1260
Qy 1261 QPPVFPVFPVGVQVPHNPSHPMTSSNHQITQCSSECDSSSHRQQLQEEHGS 1320
Db 1261 QPPVFPVFPVGVQVPHNPSHPMTSSNHQITQCSSECDSSSHRQQLQEEHGS 1320
Qy 1321 SARQGRGHRRAVPVQPCMESENENMLAEYEQRYTSDCCSSREGDTCCSSEGLIA 1380
Db 1321 SARQGRGHRRAVPVQPCMESENENMLAEYEQRYTSDCCSSREGDTCCSSEGLIA 1380
Qy 1381 EAGEPAPRQMTAKNT 1395
Db 1381 EAGEPAPRQMTAKNT 1395

RESULT 2

Q9W213

ID Q9W213 PRELIMINARY; PRY: 1395 AA.

AC Q9W213;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ROBO PROTEIN.

GN ROBO.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

CC Ephyridiidae; Erosophilidae; Drosophila.

CX NCBI_taxid=7227;

RN [1]

RF SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Annarates P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George A.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Tandell M.D., Zhang C.Q., Chen L.X.,

RA Burdon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov D.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jallal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M., McPherson D.,

RA Merkoul G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zhang X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195 (2000).
 DR EMBL: AB003458; RAP46887.1; .
 DR HSSP: P56276; 1TLX.
 DR FLTBASE: F9g0005631; robo.
 DR INTERPRO: IP0001777; .
 DR INTERPRO: IP0003006; .
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; Ig; 5.
 DR PRINTS: PR00014; FMYPEI1.
 SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13FO CRC64;

Query Match 99.7%; Score 7407; DB 5; Length 1395;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1391; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

1 MHPHPENHAIARSTSTNNPSSRSRSMWLLPAWLLVLSVAGLPAVGGQISPRIE 60
 1 MHPHPENHAIARSTSTNNPSSRSRSMWLLPAWLLVLSVAGLPAVGGQISPRIE 60
 61 HPTDVLKKNPATLCKVKGKPEPTIENKDGSPVTEKKSHEVQKDGALFRTMQ 120
 61 HPTDVLKKNPATLCKVKGKPEPTIENKDGSPVTEKKSHEVQKDGALFRTMQ 120
 121 GKRGDGGTVCANRVGQVSRHASLIQIARLDQFVPEKDFRVANGETALLEGSPK 180
 121 GKRGDGGTVCANRVGQVSRHASLIQIARLDQFVPEKDFRVANGETALLEGSPK 180
 181 GIPEPTLWIKDGVPLDLKAMSGASSRVIRVDSGNLISNVEPIDGNYKICANLNG 240
 181 GIPEPTLWIKDGVPLDLKAMSGASSRVIRVDSGNLISNVEPIDGNYKICANLNG 240
 241 TRESSAKLIVQVPMKPEKQDQVLMGQATNPHCSVGDDPPVLMKKEGNI PVSR 300
 241 TRESSAKLIVQVPMKPEKQDQVLMGQATNPHCSVGDDPPVLMKKEGNI PVSR 300
 301 RILDEKSLISNITPTDSTGYCEAHNVGQISARASLIVHAPNPTKPSNKKVGLNG 360
 301 RILDEKSLISNITPTDSTGYCEAHNVGQISARASLIVHAPNPTKPSNKKVGLNG 360
 361 VQLPCMASGNPPSVFWIKEGYSTLMPNSSSGRQVVAADGTILQITDVGQEDGTYVCS 420
 361 VQLPCMASGNPPSVFWIKEGYSTLMPNSSSGRQVVAADGTILQITDVGQEDGTYVCS 420
 421 AFSVDSGSTVRFVLSVSDVERPPPIIQGANLPKGSVATLCRATGNPSPIKNWF 480
 421 AFSVDSGSTVRFVLSVSDVERPPPIIQGANLPKGSVATLCRATGNPSPIKNWF 480
 481 DGHAVQAGNRSIIGQSSLRVDQLSDSGTTCYASGERGETSAAALTVKPSGSLH 540
 481 DGHAVQAGNRSIIGQSSLRVDQLSDSGTTCYASGERGETSAAALTVKPSGSLH 540
 541 RAADPTSPAPPTPKVLNVRSTISLWAKSQKPGAGVGIIGTVEYFSPQLGTGIV 600
 541 RAADPTSPAPPTPKVLNVRSTISLWAKSQKPGAGVGIIGTVEYFSPQLGTGIV 600
 601 AARVGDQTVTISGLTGTSTVFLVRAENTQGISVPSGLSNVIRTIADDAASANDLSA 660
 601 AARVGDQTVTISGLTGTSTVFLVRAENTQGISVPSGLSNVIRTIADDAASANDLSA 660
 661 ARTLLQKGSVELIDASINASAVRLWMLHSADEKTVBGLRIHYKDAVSPSAQHSITV 720
 661 ARTLLQKGSVELIDASINASAVRLWMLHSADEKTVBGLRIHYKDAVSPSAQHSITV 720
 721 MDAESAPFVGNLAKYTKYEFLLPFETIIGQPSNKTALTYEDVSPAPPOIQGMIN 780
 721 MDAESAPFVGNLAKYTKYEFLLPFETIIGQPSNKTALTYEDVSPAPPOIQGMIN 780
 781 QTAGVWMTTPPSQHHNGNLGYKIEVSAGNTMKVLNMTLNATTSVLLNLTGAVTS 840
 781 QTAGVWMTTPPSQHHNGNLGYKIEVSAGNTMKVLNMTLNATTSVLLNLTGAVTS 840

QY 841 VRLNSTKTAGOQPTSKPISLMDPHTHVHPRAHPSCGTDHGRHSGDLTHNMGIPPGD 900
 Db 841 VRLNSTKTAGOQPTSKPISLMDPHTHVHPRAHPSCGTDHGRHSGDLTHNMGIPPGD 900
 QY 901 INPTTHKTKTDLGSPWLMVLCVILVLVLSAISMVYFKRKQMKELGLHSVSDNE 960
 Db 901 INPTTHKTKTDLGSPWLMVLCVILVLVLSAISMVYFKRKQMKELGLHSVSDNE 960
 QY 961 ITALINKSKSLWIDHHRGWRTADTKDLSGSEKSLSHVNSSQNTNSDGGTDTAEVD 1020
 Db 961 ITALINKSKSLWIDHHRGWRTADTKDLSGSEKSLSHVNSSQNTNSDGGTDTAEVD 1020
 QY 1021 TRLNLTFTNCRKSPDPTPIATTMIIGTSSSETCTKTTISADKDSGTSTPSYDAFAGOV 1080
 Db 1021 TRLNLTFTNCRKSPDPTPIATTMIIGTSSSETCTKTTISADKDSGTSTPSYDAFAGOV 1080
 QY 1081 PAVPVKSNYLQIVPEPIWSEFLPPPEHPHPPSSITYAGQSPSSSKSSKSGSGIST 1140
 Db 1081 PAVPVKSNYLQIVPEPIWSEFLPPPEHPHPPSSITYAGQSPSSSKSSKSGSGIST 1140
 QY 1141 NQSLINASHSSSSGGTSAGWSQYAVACPPENWTSNPLSAVAGGTQWYQITPTNHP 1200
 Db 1141 NQSLINASHSSSSGGTSAGWSQYAVACPPENWTSNPLSAVAGGTQWYQITPTNHP 1200
 QY 1201 PQLPATYATPGGAVPPLMFLPFAQRAESAQGLMAACACGACRACNCDALATPSM 1260
 Db 1201 PQLPATYATPGGAVPPLMFLPFAQRAESAQGLMAACACGACRACNCDALATPSM 1260
 QY 1261 QPPPPVVFVPEWYQVHPHNSHMHPTSSNHIQVCSSECSHSSSSSHKRLQLEHGS 1320
 Db 1261 QPPPPVVFVPEWYQVHPHNSHMHPTSSNHIQVCSSECSHSSSSSHKRLQLEHGS 1320
 QY 1321 SAQRGGRHRAAPVFPQCHSENEWMALEZBQRTSDCCNSRBDGTCSCSDGSLTA 1380
 Db 1321 SAQRGGRHRAAPVFPQCHSENEWMALEZBQRTSDCCNSRBDGTCSCSDGSLTA 1380
 QY 1381 EAGEPAPQMTAKNT 1395
 Db 1381 EAGEPAPQMTAKNT 1395

RESULT 3
 ID Q9VP26 PRELIMINARY; PRT; 859 AA.
 AC Q9VP26;
 DT 01-MAR-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE C52423 PROTEIN (FRAGMENT).
 GN C52423.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celsiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Gordon R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.S.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.E., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 Ra de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Qy 170 ETALBCCPPGPPGPEPTLWIDGVLDDKAMSPGASSVRIVDGNLLISNVEPIDBG 229
 Db 148 EMALVLECSPPAGFPVWNRKDD-----KELRIQDMRTYLSADGNLLIDVDSRDSG 201

Qy 230 NYKICIAQLNWITRESSTAKLIVQVRFPMKEPKQDVMYLGQATFHCSSGGDPPPKVLNK 289
 Db 202 TIQCANNWGGERSVNFARLSYFERKPEQPKDMVDVGAALVDFCRVGDQPPQITWK 261

Qy 290 KEEGNIPVSRARILHDEKSLISNITPTDGTGYCAHNWVQISARASLIVHAPPNPKT 349
 Db 262 RKNPEPTIATYADNGLRIERVQSDSEGETYCYARNPAGTLEASGLRVQAPSPQT 321

Qy 350 RPSNKKVGLNGVQLPCMASGNPPPSVFWTKGVSTLMFPN---SSRGQTVAAQDTLQIT 407
 Db 322 KPAQDSVPAGGTATFECLVGPSPATFVSKGQDQLLPSTVSADGRTKVSPTGLTIE 381

Qy 408 DVEQDEBGTYSVCSFVDSSTVRVFLVS-----SVDERPPPIIQIGPANTQILPKGSV 461
 Db 382 EYGVQDVGATVACGNMAGSSGLSKAALKVTKATVGTNTPAKPPPIIEBGRQNTQILNVS 441

Qy 462 ATLPCRAITGNPSPIKWFHDSHVAQ-AGNRSYIQQSSRLVDDQLSSTYTCTASGER 520
 Db 442 AILPCQASGKPPGICSNLWRLGDLIDITDSRISQHSGLSHLADLKKQGVYCTIAKMD 501

Qy 521 GETSWAAILTVKEPKGST-LHRAADPSITPAPPGTKVLNVRTSILSNKASQKPAV 579
 Db 502 GEGTSWASLTVDEHTSMAQFVRMPDPSNFPSSPTQPIVTVVDTVEVLW---NAPSTSGA 559

Qy 580 PLIGITVETVSPDLQGTWIVAHRVGDQTVISGLTPTSTVFLVRAENTQISVPSGL 639
 Db 560 PGTITIIQITVSPDLQGTWENIPDYVASTEYRIKGLKPSHSMFVIAENKEIGTIPSVS 619

Qy 640 SNVTKITADDAASAN----DLAARTLLTGS-VELDASAINASVLEMLNLYSAD 694
 Db 620 SALVTSKPAQVALSDKNKMDMAIEKRLTSEQLKLEEVKTIINSTAVLFLWKKRL- 677

Qy 695 EKYVELGRILH-DASVPSAQHSITVWDASAESFVNGMLKXYTKVEFLPTE---PETI 750
 Db 678 EELIDGYIYIKWGGPPRTNDNQ---VNVTSPTSTVYVSLMPPPTVEFTYIPHSRGVSI 735

Qy 751 BGQPSKSKALTIEDVPSAPPDNIQIMNQTAGVWVWPPPSCHRNGLNKGKIEVSAG 810
 Db 736 HGAPSNMVLVIAEAPSLPPEDVIRMLMLTLIRISWAPKADGIGLILAGQI-VIVG 794

Qy 811 WTKVLANMLNATITSVLLNMLTGAIVSYRLNFTKAGDGPYKESILFMO-PTHRHV 869
 Db 795 QAPNMMNMTITNEARASVILFLVLTGMTKIRVARSNGVGVSGEYSEVIMQDLEKH 854

Qy 870 PPRARPGSGTHDREGBQDLYH---NNGNIPPEDINPTTKRTDITSLGPMWMLVCIWL 927
 Db 855 LA-----AQQENBSFLYGLINKSHVP-----VIVVAIL 883

Qy 928 VLVISAISMVYFKRKHQMKELGHSVSDNITAININSKSLW-----IDHRGKWT 982
 Db 884 IIFVILIIAYCWNERNRSGDKRSFINDGSHVMSNN---LMDVQAPNPNQPMNIT 939

Qy 983 ADTKDGLCSLEKSLSHVNSGSSNTNND--GGT-----DYARY---DTANLITF 1027
 Db 940 AGKMTNMMNNGQALYSITPNAQDVPNCCDYSMTMRPGSEHRYTALQGGGNMAMTF 999

Qy 1028 YNCRKSPDNPIYATITMIGTSSSTCTKITSISADKDSGTRSTPSDAFGQVPAVVPYK 1087
 Db 1000 TG-QTCHDPSFYATITLV-----LSMQQPA-WLN 1027

Qy 1088 SNTLQYVPEPINWSEFLPPPPHPPSPSYGYAGSPSSSRKSKSAGSISITQSLINA 1147
 Db 1028 DKMLRAPAMPTI-----PVPPP-PPARTADHTAG-RSSRSRSDAGRC-----TLNG 1072

Qy 1148 SIHSSSGGFWAGVSGYQAVACPPENYVSNPLSAVAGGTQNTQITPTIMHPQPL- 1203
 Db 1072 GLHRTSGSQBS-----DSPPHTVSGVQLHSDSGTGSSEKRTGEERTPPKXLM 1123

Qy 1204 ----PATFATITGGGAVPNHLPFATQIRHAASEYQAGLNARCAQSRACNCDALATPSM 1260

Db 1124 FIPPPSPNPPPGGVYDIDFQATRAQ-----LNRGSPREDTYSG----- 1165

Qy 1261 QPPPPVPEGVQVPHNPNHMTPTSNRQITQSSRCSHSRSQSKRQLQLEHGS 1320
 Db 1166 -----VSDGAFARVDVNA---RPTSRNRNL---GGRPLAGKRDQSSRLNMDODGG 1212

Qy 1321 SAQRQGGHRRAP 1334
 Db 1213 SSRADGENSGDVP 1226

RESULT 5
 089026
 ID 089026 PRELIMINARY; PRT: 1612 AA.
 AC 089026;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE DUT1 PROTEIN.
 GN ROBO1 OR DUT1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Wu M.C., Lowe N., Fordham R., Rabbitts P.;
 RT "The mouse homologue of human DUT1/H-robo1 gene: protein sequence and
 chromosomal location."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L71793; CAAT:6850.1; -
 DR HSP: P56276; LTK.
 DR MGD: MGI:124781; Robo1.
 DR INTERPRO: IPR001777; -
 DR INTERPRO: IPR003006; -
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; Ig; 5.
 SQ SEQUENCE 1612 AA; 176406 MW; 572988544796848 CRC64;

Query Match 21.6%; Score 1607.5; DB 11; Length 1612;
 Best Local Similarity 30.9%; Pred. No. 1.5e-99;
 Matches 426; Conservative 189; Mismatches 485; Indels 277; Gaps 40;

Qy 56 PRIIEHPDLVKKNEPATLNKVGKEPTEIWFKDPGEVST---NEKXSHRVQKDGAL 113
 Db 29 PRIVEHPDLVSKGEPATLNKAGRPPTPIETWFGGSRVETDKDPSHRLMPSGSL 95

Qy 114 FFYRTNGKKKKQ-DGGETVCVAKNRVQAVSRHASLQIALVLRDQFVEPKDTRVANGETA 172
 Db 89 FFLIRVHGKSRDQPGVYICVARNLGVASHNASLEVALRDQFQNFSDVMVWAGSPA 148

Qy 173 LLECGPPGKPEPTLWIDGVLDDKAMSPGASSVRIVDGNLLISNVEPIDBGNYK 232
 Db 149 VMCQPPGHPPEPTISWKKDGLD-----KDERITI-AGGLIMITTKSDAGKYV 200

Qy 233 CIAQLNWITRESSTAKLIVQVRFPMKEPKQDVMYLGQATFHCSSGGDPPPKVLNKK 252
 Db 201 CVGTNMGVRESSEVARTLVLERPSFYKRSNLAIVDVSASFCKAARGDPPVTVMKDD 260

Qy 293 GNIPVSRARILHDEKSLISNITPTDGTGYCAHNWVQISARASLIVHAPPNPKRPS 352
 Db 261 GELPKSRTEI-RODHTLKRVTAGDMSGYTCVARNMGKARASATLVQEPPEFVVKPR 319

Qy 353 NKKVGLNGVQLPCMASGNPPPSVFWTKGVSTLMFPN---PNSSBGRQTVAAQDTLQITV 409
 Db 320 DQVWALGRVTFQCCATGNMPQPAIWRREGSKNLLSTPQSSSRFSVSGQDGLITNV 379

Qy 410 RQDEBGTYSVCSFVDSSTVRVFLVS---DERPPPIIQIGPANTQILPKGSVALPCRA 468
 Db 380 QRSVDGYTCITQILNAGSITIKRVELVDIARDPPVPIRGVQVNTVADGTLISCA 439

Qy 469 TGNPSRIKWFHGDHAWQA-GNRTSIQGSGLVDDQLQSDSGTTCASGERGETSWAA 527
 Db 440 TGSPPTLILMKRQGLVSTQDSRIKQLENSVLQIRAKLQDGTGTCTASTPSGEATWSA 499
 Qy 528 TLTVKEKP-STSLHRAADPSTYPAPGTPKVLNVSRTSISLWAKSQEKPGAVGPIIGT 586
 Db 500 YIEVQGFVQVPPRPPTDPLNLPSPASKPEVTVDSKNTVLWS--QPNLNSGATPTSTI 556
 Qy 587 VEYFSPDLQGVIAVAHRVGDQTVTISGLTPTGTVSYVLVAENTQGISVPSGLSNVITK 646
 Db 557 IEAFPSHAGSGSQVTAENVKTTETPAIKGLPNAILPLVRAANAYGISDPSQISDPVKQ 616
 Qy 647 EADPDASANDLSAARTLLTCKSV-ELIDASAINASAVLEWMLVSADEKTVGLRIHY 705
 Db 617 DVFPPTSGVQVHQVREL--GNVVLHNPILNLSSESVHVT--VDQSQVTIQKILY 672
 Qy 706 K--DASVPSAQVHSITVMDASAESVVGWLNKTKYEFLLTPFTTIBGQPSNKSATLY 763
 Db 673 RPSGASGSESLNLFVFTPTKNSVVPDLARKGVNIAKRPFFKPGQADSKPARKYL 732
 764 EDVPSAPPNOIGMT--NQTACWYRWTPPSSQHNNKLYGKIKVNSGMYNMLNANL 821
 Db 733 EAFSPAPPSVTSKNDGNSTAILVTWQPPPEQVQMGVQEKV-HCLGNETRYHNTKV 791
 Qy 822 NHTTISVLLNLTGAVTSVRLNSPTKACDGPYSKISLMDPTPHVHPRAAPSTHDC 881
 Db 792 DGSFTSVIPLSLVFGIRTSVEAASGTAGSPGKSEPFOLD----- 833
 Qy 882 RHESQDLTYHNNGN-IPGPDINTPTHKTTDYLSP-----MLWMLCVIWLVLV 930
 Db 834 -----SGSNFVSPED-QVSLAQGISDVWQPAFIAGTACMCLILNFSWL----- 879
 Qy 931 ISAAISMWYFKRKHQ-----WTKELHLSVSDNEITALNINSK 969
 Db 880 -----YRHKKNGLTSTYAGIRKVPSTPTPTVTYVQGGVAGSGGPGGLNISEP 931
 Qy 970 ESL-WI-----DHRGWRITADTKDGLSEKLLSHVNSQ--SNTNNS----- 1010
 Db 932 ATQMLADTWPNTGNNHDCSINCTAGNGNSDNLTYSRPADCIANTNQLNQNKL 991
 Qy 1011 --DGGTQAEVDTANLTATFYCNKRSKD-----NPTPIATMII----- 1046
 Db 992 MLPESTVIGDVLSSNINDMKTFSNPLKDRFVNSFGQPTPYATQLIQANLSNNGN 1051
 Qy 1047 -GTSS-----SETCTATISADKDSGTHSPY 1072
 Db 1052 AGDSSEKHNKPPGQKPEVAPIQYIMQNKLNKYRANDTPTPIPTNQSTQNTGSSY 1111
 Qy 1073 SDAFAGVPAVPVYSKNTLQYVPEP-----INNSLEPPPEHPSSSTYGTAGSPSSR 1128
 Db 1112 NSSDGSSTSGSGCHKKGARTPKAPQCGKNWALLPPAPPPHS----- 1158
 Qy 1129 KSKSGAGSISTQISILNASISGSSGGFSAWGSQVYAVACPENVTSNPLSAVAGTQ 1188
 Db 1159 -----NSEETNMSVDES-----YDQEMCPVPPAPMTLQ-----Q 1188
 Qy 1189 NRQITPTNPBPPLPAYFATTGPGGAVPNHLFPATORHASEYQAGLNAARCASQAC 1248
 Db 1189 DEIQ-EEEDERGGTPPPVGAASSP-AAVSYSHQSTAN----- 1223
 Qy 1249 NSCDALATPSP--MQP-----PPPVVPEGQTVQVPHNSHPMHTSSNH 1290
 Db 1224 -----LTPSPQELQQLMLOCPEDLGHMHP--PDRRQVPSV-PPPPRISPPH 1270

DE ROUNDABOUT 1.
 GN ROBOT.
 OS Homo sapiens (Human).
 OC Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98117249; PubMed=9458045;
 RA Kidd T., Brose K., Mitchell K.J., Petter R.D., Tessier-Lavigne M.,
 RA Goodman C.S., Bear G.;
 RT "Roundabout controls axon crossing of the CNS midline and defines a
 RT novel subfamily of evolutionarily conserved guidance receptors".
 RL Cell 92:205-215(1998).
 DR EMBL: AF040390; AAC39575.1; .
 DR HSSP: P56276; 1TLK.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; dg; 5.
 SQ SEQUENCE 1651 AA; 180928 MW; 98D987CAB73074D CRC64;

Query Match 21.44; Score 1592; DB 4; Length 1651;
 Best Local Similarity 30.24; Ref. No. 1.7e-98;
 Matches 419; Conservative 188; Mismatches 492; Indels 290; Gaps 39;

Qy 56 PRIIEPTDLVKNNEPATLNCKYEGKPEPTIEWFKDGEVST--NEKSHRVQFQDGL 113
 Db 68 PRIVEHSDLIVKSGEPATLNCKAGRPPTIEWYKGEKVEKTDKDPASHMLPLSGSL 127
 Qy 114 PFTRMQKKKQ-DGGEYWCVAKRVGQVSRHSALQIAYLRDDFVPEKDKRVANGETA 172
 Db 128 FFLRIVHSKRPDGEVTVCAVNLKGVASHNLSLEVALRCDPRNPSDVWVANGEP 187
 Qy 173 LLECCPPGIEPTLILNKGVDLIDKAMSGASSRVIRVGGNLLISNVEPDNSNY 232
 Db 188 VMEQPPPHPEPTISMKKDGSLDD-----KDERITI-RGKLMITTKRSDAGKYV 239
 Qy 233 CIAQMLVTRDSYSKALIVQVRFKPKPKDQVMLQYATFHCVSQDGPFFVKMLKEE 292
 Db 240 CWTNMGVSESEVALTVLEKPSFVKRSLNLAIVDLSAEFKCAERGQVPTVWRKDD 299
 Qy 293 GNIPVSRARILHDEKLSNITPTDGTGVRANHWQGISARASLIVHAPMTWKPS 352
 Db 300 GELPKSRILE-KDHTLKLKRVAGSDMGSTVADZWGASASATLIVQEPHFVFKPS 358
 Qy 353 NKKVGLNVLVQLPMAASNPSPSVFVTKGVSTLME--PMSHGKQVTAADQLITVD 409
 Db 359 DQVVALGKVTFOCEATGNPQAFIRKREGSKLLSTPQPPSSSRKFSVSGQDLITNW 418
 Qy 410 RQEDGTYTCASVSDSSTVRVFLQVSSV-DERRPQITGIPNQTLPGSVATLPCRA 468
 Db 419 QRSVDQITQLNAGSIIITKALEVDTADRPVPPVIRGPNQVAVDGTFLVSCVA 478
 Qy 469 TGNPSRIKWFHGDHAWQA-GNRTSIQGSGLVDDQLQSDSGTTCASGERGETSWAA 527
 Db 479 TGSPVPTLILMKRQGLVSTQDSRIKQLENSVLQIRAKLQDGTGTCTASTPSGEATWSA 538
 Qy 528 TLTVKEKP-STSLHRAADPSTYPAPGTPKVLNVSRTSISLWAKSQEKPGAVGPIIGT 586
 Db 539 YIEVQGFVQVPPRPPTDPLNLPSPASKPEVTVDSKNTVLWS--QPNLNSGATPTSTI 556
 Qy 587 VEYFSPDLQGVIAVAHRVGDQTVTISGLTPTGTVSYVLVAENTQGISVPSGLSNVITK 646
 Db 596 IEAFPSHAGSGSQVTAENVKTTETPAIKGLPNAILPLVRAANAYGISDPSQISDPVKQ 655
 Qy 647 EADPDASANDLSAARTLLTCKSV-ELIDASAINASAVLEWMLVSADEKTVGLRIHY 706
 Db 656 DV-LPFSQGVHQVQRELGNVNLHNPVLSSSESVHVT--VDQSQVTIQKILY 712
 Qy 707 DASVPSAQVHSITVMDASAESVVGWLNKTKYEFLLTPFTTIBGQPSNKSATLY 760
 Db 733 EAFSPAPPSVTSKNDGNSTAILVTWQPPPEQVQMGVQEKV-HCLGNETRYHNTKV 791

RESULT 6
 Q976N7

ID Q976N7 PRELIMINARY; PRT: 1651 AA.
 AC Q976N7;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

Db 713 ---PGSAGNCSDLVFEVTPAKNSVPIPLRGKVNTEIKARPPNFEQAGDEIKFA 768
 Qy 761 LTIEDVAPPONDIQIGMY--NOTAGVWRMTPPSOHHNGLGKIEVSAGNTMKVLN 818
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 769 KILZEPASAPPQGVTVSKNDGNTALVLSWQPPEDTQNGVQRTYV-WCLGNETRYEIN 827
 Qy 819 MTLNATLTIVLLNLTGAVTSYRLNSFTKADGSPKPSISLMDPTTHVHPFAPRPSGT 878
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 828 KTVGSGTFSVYFPLVPGIRTSVEAASGTAGSGVSESPQIQLD-----AH---- 874
 Qy 879 HDGRHGQCOLTYNNGN-IPPGDINPTHKTKTDYLSGPP-----WMLVNCIVLL 927
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 875 -----GNPVSPED-QVSLAQISDVVQKPTAGIAGACWILLMVSFSLW 918
 Qy 928 VLVISASIVTSYFKRKHQ-----MTKELGHLSTVSDNEITALNI 966
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 919 -----YRRKKRNGLTSTYAGIRKVPSPFTPTVTITQRGGAENVSSGGRGLLNI 967
 967 NKSLEL-WI-----DHRGWRTADTDKSDGSEKLLSHVNSQO-----SNYNS----- 1010
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 968 SEPAAQWILADTWPTGNNHNDSCISCTAGNGNSDLNLTYSRACDIANYNQNLNKO 1027
 Qy 1011 ----DGGTDAEVDTRNLNLTYPNCRKSPD-----NPTPTA-LNIGTSSETC 1054
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 1028 TNLMLPESTVYGVDLSKNINEMKTFNSPKWKGDFVNPSCGPTPTATLQILNSLNNM 1087
 Qy 1055 KTKTTSIAOK-----DSGTH 1059
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 1088 NNGSGDSGEKHKWQKQZVAPVQYNIQVNLKNDYTRANDVPTPTIPYNTQMTG 1147
 Qy 1070 SPYSDAFAGQVPAVPVYSWLYQYFVE-----TNWSEFLPPPEHPPPSSITYAGQSP 1125
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 1148 GSTNSDGSSTSSQGHKKGATPKVKQGGNWADLLPPAPHPHPS----- 1197
 Qy 1126 SSRKSKSAGSGISTNQSLNASISBSGSGFSWAGVSGQVAVACPNVNT----- 1176
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 1198 -----NSEEYNISVDES-----YDQEMCPVPTPARYLQDLEEE 1233
 Qy 1177 -----SNPLSAGVAGTQNTYQITPTINQH--PPPLPAFATPGGAVFPNH 1220
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 1234 EDERGPTTPVPRKASFE-ANVSYSHQSTATLPTSQEELQPMQDCEPTEG-----H 1284
 Qy 1221 LPATQGRHAESQAGLNAARCAQSRACNSCALATSPMQPPPPVVPVSGWGTQVPHNS 1280
 Db 1285 MQHQPKRRK-----QPVSPPPP-PR-----ISPHTYG 1312
 1281 HPMPPTSSN 1289
 Db 1313 FISGLPVSQ 1321

RESULT 7
 05005

ID 0505005 PRELIMINARY; PRN: 1651 AA.
 AC 0505005
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE TRANSMEMBRANE RECEPTOR ROBO1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPINAL CORD;
 RX MEDLINE=98117249; PubMed=9458045;
 RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
 Goodman C.S., Tear G.;
 RT "Roundabout controls axon crossing of the CNS midline and defines a
 novel subfamily of evolutionarily conserved guidance receptors.";
 RL Cell 92:205-215(1998).
 DR EMBL: AF041082; AAC39960.1; .

DR HSP: P56276; 1TLK.
 DR INTERPRO: IPR001777;
 DR INTERPRO: IPR003006;
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; ig; 5.
 KW Transmembrane.
 SQ SEQUENCE 1651 AA; 180746 MW; FA2452D046E18687 CRC64;

Query Match 21.3%; Score 1585; DB 11; Length 1651;
 Best Local Similarity 29.0%; Pred. No. 5.1e-98;
 Matches 456; Conservative 207; Mismatches 538; Indels 374; Gaps 49.

Qy 44 NGLPA-----VRGQYSPRIIEHPTDLVKNKNPATLCKVGVKPEPTI 87
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 40 NGTPAPTSNDNDNSLGTGSRLRQEDPPRIVEHPSDLVSKGEPATLCKAAGRPPTPT 95
 Qy 88 EWFKDGPEVST--NEKSHRVQKDGALFFYRTQKQKQD--DGGETVCWKNRVQVSR 164
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 100 EWFKQGEVETDKDQPRSHMLPGSGSLFFLRVHGKSRSPDEGVTCVARMVLGVSH 159
 Qy 145 HASLQIAVLQDREVEPKDTRVAKETALLRCGPPKGPPEPTLWIKQGLVDLAKMSF 204
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 160 WASLEVALLRDQPNQSDVMVAVGEPAWECOPRGHPPEPTISWKGQSLD----- 213
 Qy 205 GASSRVRYDGNLLSNVEPDEBGNVACIQNLVGTRESSYARLIVQVFPYMEPKQD 264
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 214 -KDERITI-RGKMLITITRKSDAGKYVCTGVNMGERSKVDVTVLERSPVIRPSNL 271
 Qy 265 VMYQQTATPFCSSVGGQPPKVLWKEBGNFVPSARILLHDEKSLKISNITPDGQITVC 324
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 272 AVTVDQSAEFCARGDPPVPPGWKDGDELPSKYEI-RDHTLAKIKVIAAGSDGTC 330
 Qy 325 EARNWQGISARSLIVHAPMTTRKSNKVLGNVGLPCMASGNPPFSVFWTEGVS 384
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 331 VASNMGKASASLIVQEPFVFPKQDQVVLGKRVITPQCEATGNQPAIWAEGSQ 390
 Qy 385 TLMF---PNSSHGQIVNADGLQITDVRQEDGTVCSAFVSQDSSTVWVLQVSSV 440
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 391 MLSPSTPQSSSSRSVSQDGLTWTWQKSDGVTYQILNVAAGSITATLEVDYTA 450
 Qy 441 ERPPFIQIGFAMTQVSGATLPCRAITGNSPRIKWFHGHQAQV--GNRTISQGSLL 499
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 451 DRPPFVIRGQVQIVADGTLTLSCVATGSPVPTLWKKDGLVSTQDSRIQLGSEVL 510
 Qy 500 RVDDLQSLDSSTYCTASGERGTSAATLIVKPK--STLSLRADSTPTAPPQTPKVL 558
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 511 QIRYAKLGRTYCTASTPSGATGSAVIVQGEFVVPVQPRFTDNLIPSAKSPKVT 570
 Qy 559 NVSRITSISLRKSKQKPGVGPVIGITVEYFSDLQGWIAVHRGDTQVITSLGTP 618
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 571 DVSNKNTLLW---QPNLSGATPTSYIIBAFSHAGSGSWQVARNYKTETFAIGLKN 627
 Qy 619 TSTVELVRAENTQGISVPSGLSNVKITRADPDASANDSAARTLLTGKSV-ELIDASA 677
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 628 AIYFLVRAAMATGISDPSQISDPVTKDQVPTTQVDHKGQVREL--GNVYLNHPTI 685
 Qy 678 INASAVLEWNLVSADEKYVGLRIHLY--DASVPSAQHSITVMDASAESFVGNLKK 735
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 686 LSSSSVEHWNT--VDQSQYIQGYKILRPSGASHGSEDLVFEVTPAKNSVPIPLR 743
 Qy 736 YTYEFLIPFETIEGQPSNSKATLYVEDVSPAPNDIQIGMY--NOTAGVWRMTPPPS 793
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 744 GWTEVLEKARPPNFEQAGDEIKFAKTLLEERSAPPSPVTSKNGCNGTALLVQPPPE 803
 Qy 794 QHNGNLTGYAIEVSAGNTMKVLAMTNATLTIVLLNLTGAVTSYRLNSFTKADGSP 853
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 804 DTQNGVQRTYV-WCLGNETRYEINHTKMTVDGSESTVPLVPGIRTSVEAASGTAGSG 862
 Qy 854 YSKPISLMDPTTHVHPFAPRPSGTHDGRHGQDLTYNNGN-IPPGDINPTHKTKTDY 912
 ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||: ||:|||||:
 Db 863 KSEPPQIQLD-----SHQNPVSPED-QVSLAQISDV 893

QY 913 LSGP-----WLMVLVCLVLLVLSAISMYVFRKKQ----- 945
 Db 894 VQKQAFIAGIACACIIMLVFISWL-----YRHRKKNGLSSTYAGINRVSPT 942
 QY 946-----MYKELGHSVYSNEITALNINSKESL-WIDHWRG-RTADTKD-----SG 990
 Db 943 PTPVTYVQRGGEAVSSGGRGLNLSSEPATQWPLAD--TWPTNGSHNDCINCCXTASG 1000
 QY 991 LSESKLLSHVNSQ--SNYNS-----DGGTDYAEVDTNRLTYVNRKSKPD----- 1035
 Db 1001 NDSNLTTSRPAQCIANYNQLDNKKQTNLMLPESTYGVDLVSKNEMKNTFNSMLKD 1060
 QY 1036-----NPTPIATTMI-----GTSSSE----- 1052
 Db 1061 GRFVPSGQPTPIATTQLQANLINNMWGGDSSEKHWKPGQCEVAPIQYINMQN 1120
 QY 1053-----TCTXTTSSADKSQTHSPSDAPAGQVPAVPPVKSMTLQVPEP-----I 1098
 Db 1121 KLMKDTRANDTLPTIPYHSTYDQNTGGSYNSDRGSSCTSGSQGHKARTPKAPKQGM 1180
 QY 1099 NWSEFLPPPPPPPPSSST-----YGYAGSP-----ESSRKS 1132
 Db 1181 NWADLLPPPPAPPPPHNSSEYSSMSVDSYDQEMCPVPVPMYTLQDEEEAEGRPT 1240
 QY 1133 SAGSGISTQYSLNARSISHSSSGGFSANGVSPQYAVA-----CPENYVSNPLSAVAGT 1187
 Db 1241 PPVKAASPAANSYS-HQST-----ATLTPSQEELQWMLQDCPD-----LGMHWP 1289
 QY 1188 QNRVYITPTNQHPPLP-----AYPATG----- 1212
 Db 1290 DRRQ--PWSPPPPPIPSPTTYGISQVNSMDTDAPEEEDAEQWAKMQTRLL 1347
 QY 1213-----GGVPPNHL-----PPATQR 1227
 Db 1348 LRGLQETPASSGQDLESSVTGSMINGSSASEDNISSGRSVSSSDGFFTDADPAQV 1407
 QY 1228 HAASETQAGLAAARCAQSRACNSCDALATPSMQPPPPVPEGHTQVPMNSHPHPTS 1287
 Db 1408 AAAAEY-AGLVARRQKQDAAGRRHFASQCP-RTSPVS-----STD 1447
 QY 1288 SNWQITQSCSSDHSRSSQHKRQLLEERGSSAKRQGHRRRA-----PVQPC 1339
 Db 1448 SN-----MSAAVIQARPTKQKH-----PGLHREAYTDOLPPPPVPPA 1489
 QY 1340 MESENNENLAETQR 1354
 Db 1490 IKSPVQSGAKLEAR 1504

RESULT 8
 Q213

QY Q9Q213 PRELIMINARY; PRG: 1060 AA.
 AC Q9Q213;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxID:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99200391; PubMed-10102268;
 RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
 RA Tessier-Lavigne M., Kidd T.;
 RT "Slit proteins bind Robo receptors and have an evolutionarily
 RT conserved role in repulsive axon guidance.";
 RL Cell 96:795-806(1999).
 DR EMBL: AF182037; AAF0458.1; .
 DR HSSP: P56276; ITLK.
 DR INTERPRO: IPR001547; .
 DR INTERPRO: IPR001777; .

DR INTERPRO: IPR003006; .
 DR PFAM: PF000041; fn3; 3.
 DR PFAM: PF000047; ig; 5.
 DR PRINTS: PR00014; FNTYPEII.
 DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
 KW Receptor.
 FT NON_TER 1060 1060
 SQ SEQUENCE 1060 AA; 116790 MW; C48C8C118542DA CRC64;

Query Match 20.24; Score 1499; DB 11; Length 1060;
 Best Local Similarity 34.74; Pred. No. 1.7e-92;
 Matches 361; Conservative 175; Mismatches 398; Indels 106; Gaps 28;

QY 56 PRIIEPTDLVKKNEPATLCKVEGKPEPTI----EMFKDG-EPVSTNEKSHRVQKD 110
 Db 31 PKXVEQSPVIVSKGKPNTPMKQKGRFPPTIGVQVQWVKPGWKTDDSKVTQCLLPS 90
 QY 111 GALTFTYTKQKKQKQ--DGGEYCVAKNRQVQVSRHASLQIAVLRDFDFVEKDTRVAKG 169
 Db 91 GSLFLRLIYHRRSKPDGEGTVCVARNYLGCAVSRAASLEALLRDFRNPNTDVVAAG 150
 QY 170 ETALLCQPPGPIPEPTLWIKQVLDLAKMSFGASSRVIVDGNLISNVPIDEG 229
 Db 151 EPAILBCQPPGHPPTIYKKDKVIDE-----KEERISI-RGKLMISNTKSDAG 202
 QY 230 NYKICQNLVWTRSSYAKLQVQWPKMKPKQDMQLQQTATFHSVQGDPPVPMVK 289
 Db 203 MYTCVSTNMGVGRSDPAELIVFERPTLARPPIVQVLEDAEPFQCGQDQGPVPMK 262
 QY 290 KEGENIYVSTARIHDEKSLISNTPDEGTGYEABHNQVQISARASLIWHPNFTK 349
 Db 263 KDDADLPFGHDI-KDDYTLRIKAISADETGYCIENRWKGVASATILVTPAPQVW 321
 QY 350 RPSNKNGLNGLVQVLPCHASGNPPSPVFWKEVSTLMPFN-----SSHQGRVVAADTLQI 406
 Db 322 RPDQVLAQVAGTITPFCETGNQPAVFWQKESQMLLPQWQPKNSCSVSPIDGTL 381
 QY 407 TDVRQDEGTGYSAFVSVDSTVRLVQLVSSV-DERRPPIQIGANQLPKPGSVATLP 465
 Db 382 TNIGRSDAGITCQALTAVAGTIAKAGVLEVDVTDTRPPPIQLQPIVQLVADGVALTK 441
 QY 466 CRAFTGNPSRIKWDHGAUVAGNYSIIQS-SLRVLDQLSDSGTITCTASGERGTS 524
 Db 442 CKATG-PULPVIWLGKGTFLGRDPATIQDQGLTKNLRLSDGTTCVATSSGSETS 500
 QY 525 WAATLVTKPGSTLSHRAADPSTPIAPPQTPKVNLWRTSISLRWASQKPGAVGPIIG 584
 Db 501 MSALVLTGSGAT-ISKNYDTNLPQKSPQVTDVTKNSVTLQWQ-GTGV-PASA 556
 QY 585 YTVFSPDLQGVIAHVRGDTQTISGLTGSTVYVLRANTQGISVPSGLSNVIK 644
 Db 557 YIIEASQSVSNQWQVANHVKTLTLVWGLRPNITLWELWAIINPQGLSDPSPDPR 616
 QY 645 TIEADFASANDLSARTLLTKGVELIDASINASAVRLWMLHSADEXVLEGLRH 704
 Db 617 TQDIS-PPAQGVDRHQVQKLEQDVLRLNRPVPLPTTVQVWT--VDRQPTIQGVW 673
 QY 705 YKAS--VPSAQVBSITVMDASAESVPMGKLYTKYEFILPPFEITBGQPSNKTAL 762
 Db 674 YRGTSLQASTVWNLDAKPTERSAVLNM.KGVYTEKVPPIFNQMGDSEKTLRT 733
 QY 763 YEDVPSAPDNI---QIGMYNQAGVWVTPPPSQHNGNLYGKIEVSAGNTKVLAMN 819
 Db 734 TEAPASAPQSVTVTLVSGHNSISVSWDPPADNQNGIIOEYKI-WCLGNETRFHKN 792
 QY 820 TLMATTSVLLNLTTCGVSVRLNSFTKAGDGYPSKISLNDPETHVHPRAHSPGTH 879
 Db 793 TVDATISVVLGGLFGIQRVEVAESTSAGVGVKSESPDII----- 835
 QY 880 DGRHGEQDLYTHNNGIPPGDINPTHTKTYDLSGCPWMLVCLVLLVLSAISMYV 939
 Db 836 GGRNE-VVITENN-----STRTQTDVWQKQAFIAGIAGACWMLGFSI-WLY 883

RESULT 10
092214

QY	38	LVLSAAGLVL - AVRGQY - -----	SPRIIEPTDVLVKNEKPAILKCVG	61
Db	8	LTQSKPGLPPVALPGYGLTLPSSPGSRGVPEDMAPRVYQPQDVLVSRGEPATLPCRA	67	
QY	82	KPEETIWFKQGEFVST - NKKKSEFQPKDGLAFYTRTQMKKG - DGGSTYCAVNRV	138	
Db	68	RRPRNIETWKAGAVATAREDPANRHLPLPSGALFFPRIVHGRSRSPDGYTICVARNLT	127	
QY	139	QQAVSHASLQIALVLDKDFPEVKOTWAKETALLCGPGKIPPTLVIKIDGVLFD	198	
Db	128	GAASRNASLQVLAIVLDQFSGQNPVYVAGFVAFVCFKPGKIPPTLWKKGIKLKE	187	
QY	199	LKAMSPGASRWIVDGGNLLISNVEPIDEGNYKCIAQVLGTVRESSTAKLIVWKPPM	258	
Db	188	----- EGRITIT - RGGKLMSTFTKSDAGMTVCASNAGREGSAGHVLRLSP	239	
QY	259	KEPKDQVVLIGYQTATKCSVGGDPPKVLWKEEGNIPYSRATLDEKSLSEINITPD	318	
Db	240	RRPIVQVLADAFVNFCLSVGGDQPKMLHWRKDGELPAGETRESQ - SLIMDQVSSD	298	
QY	319	EGTVTCERNWVQGISARASLIVHAPNNPTKPSNKKVLGNGVQLPCASGNPPSVF	378	
Db	299	EGTITCVAGNSVGRASAGSLSVHVPQFYTKQNTYVAGNVSQCETKGNPPPIAF	358	
QY	379	TKGYSVTFMPNSS - - HGRQIVADGLITIDVRQZDEGTVCSFASVDSSTVRFVIG	435	
Db	359	QKGSQVQLFPPSQSLVGRGLIVLSPRQQLINIVKIQDGGTYVQAVSAGSLAKALLE	418	
QY	436	V - SSVDPERPPIIQIPGANTLPKGSVATLPCRTGNPSPRIKWHDGHAVQA - GNRS	492	
Db	418	IKGASIDGLPILQ - GPANQTVLWGSVWLPCRVIGNQPTQWKDKERLQDQSQFN	477	
QY	493	LIIGSSLRVDLQLSDSGYITITASGERGETSWAATLTVKPCSTSLRAADPSTYPAP	552	
Db	478	LMQNGTILASIQMDMGSPYCVAKSSIGRATNWSLWKLQEDNGASPGATPGSPNGPP	537	
QY	553	GTPKVLNYSRTSLSLWAKSCKRPGKAVYIGITVYTSFLDQITWIAAHRVDQITQY	612	
Db	538	SQPIVETVYANSTILV - KPMQSGSQA - TSYVIEAFSSQAQNTWRIVADGVQLTETI	594	
QY	613	SGLTPTGTSVTLVLAENTQGISVSPGSLNWIKITIADPDAAASANDLSAARTLLQKSV	672	
Db	595	SGLOPNTIIVLIVRAGWGLSESPSVSPVOTDSSL - SRPQKPKWGCGRLAEAVM	653	

QY 673 IDASALNSALVLEKMLHVSADKEYVEGLAHYKDAVSQAQSHVSTVMDASAEVSFVGM 732
Db 654 QEP7YLGRTIGQVSWT-----VGGPVQLVQGFVSWRIRGLDQGSWMLDLQSPHQSTVLRG 711
QY 733 LKX7TYEFYLLPFETIEQGPSNKITALYEDVPASDPNMQI-----GMNQTAGWVRPT 790
Db 712 LPPGAQIQIKVQWQGBGLGASFPVTRSIPEDABSGPQCVAVALGDGGRNSVSWMEP 771
QY 791 PPSQHBHNLNQLKYIEVSAGNTMKYLANMLNATT5VLLANLTGAVTSVRLNSPTKG 850
Db 772 PLPSQRNVIYETQI-WCLGNESRPHLNRASGAWRSV75GLLPQIYRALVAATAASG 830
QY 851 DGPYSKPLSLMDPTHRHVPRAHPSQDHDGRHEQDILTENNNGIPGDIINPTHKKT 910
Db 831 VGVASAPVQLP-----PPPAEFG-----PVSSEGLAEIRA 863
QY 911 DYLSGPMLWY-----LVCVILLVLYISAISMYYKRRKHTKGLHLSVNDNEITALIN 967
Db 864 KYLRKPAFLPAGLSAAGCALLGLPCA-----LYARQQRKELSHY-ASFA7YPAVSF 916
QY 968 KSLS-----WT-----DHRGW-----P7ADTDKSGLSSE 994
Db 917 HSEGLSGSSSRPMLGPAATYFWLSDWPPHPPSPSAGPBGSCSNPDP-DRYNEA 975
QY 995 KLSHVNSQSNNTNDSGGDTAYAVDT-----RNL7TYNCRKSPNDPTPIAT7MIIGTSSSE 1052
Db 976 GISL7LAQ7ARGANASGEBQSVYTDIFVGEEL7THFGGPOHSSGDPS7NQTAPENSE 1035
QY 1053 TCKTNTS7ADKDSG7SPSPSADAFAGVAPVVPKWSN7LQYVPE-P7INMS7FLPP7EH 1110
Db 1036 -----GDSG-----ARGG-----GK7LGPVWPSL5NPALPP----- 1065
QY 1111 PPPS7T7YAGQSPSSRKSXSGAGSIS-TWQ5ILMS7HSSSSGGFSANGVSPQVAF 1168
Db 1066 PPPSCSLGCPG-PEELGAGSDLEWCPPVPK5HL-----VGS5SGA-----CM 1111
QY 1169 ACPPE7NYSN7LNAVAGQ7NRYQ7IPT7NHPQLPATF-----ATTGCPGVP7N 1219
Db 1112 VAPAPD7FSP7ST5Y-GQ5STATILTSPSPQ7P7D7IPLH7QW7P7R7VLP7GS----- 1164
QY 1220 HLPFATQ7HASE7Q-----AG7LNARCAQ7SRACNSCALAT7SP7Q 1261
Db 1165 -----PL5YQPAL5SHDR7PVLGACQ7VLSY-----ASP5PV7TASAPRGTQV7G 1214
QY 1262 -----PPPPVPEV-----G7YQV7PHNS 1280
Db 1215 EM7P7LGH7R7AIR7KK7PAL7YR7HSGD7LPP7LP7PE7L7K7L7G7SAG7SR7V7F7RA 1274
QY 1281 HPN7PT7SNH7Q7YC7SSE7CS7HR7SG7HQ7RL7Q7LE7H7G7SAP7K7OR7GH 1328
Db 1275 RA-----CWQEE5SAGASAS-----RGT5S7ORGP7H 1299

RESULT 11
P91572
ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN 2K377.3.
OS *Caeenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caeenorhabditis.
CN NCBI_TextID=6239;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Atkinson R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copest T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.

```

RESULT 12
001632
ID 001632      PRELIMINARY;      PRT;      874 AA.
AC 001632;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA CESC12R.

```



```

RESULT 13
P97603
ID P97603 PRELIMINARY; PRT; 1377 AA.
AC P97603;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE NDOGENIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97015074; PubMed=8861902;
RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.Y.,
RA Culotti J.G., Tessier-Lavigne M.;
RT "deleted in Colorectal Cancer (DCC) encodes a netrin receptor."
RL Cell 87:175-185(1996).
DR EMBL U68726; AAA1100.1; -.
DR HSSP: P56276; 1TLX.
DR INTERPRO: IPR000531; -.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003046; -.
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; ig; 4.
DR PRINTS: PR00014; FNTPEI11.
DR PROSITE: PS00430; TOMB_DEPENDENT_REC_1; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1377 AA: 150637 MW: E514ED8AD1A63A9 CRC64:

```

[illegible]

Qy 1104 -----LPPPPHPPPSTGYAQGPS-----SKSSKSAGSGI--S 1139
 Db 905 RRSSTWSMTAGATFELVPTSPKDVTVSEKGRPTIIVNNQPPSBAKTIYGIITS 964
 Qy 1140 TQNSILNASHSSSGGFSAMGVSP-----QYAVCPENYVSNPLSA 1182
 Db 965 TD--VNAETHD-----WYIEPVGNRLTHQIELDTPTFYIKIARN--SKMGFP 1011
 Qy 1183 VAGGTQNYRTIPIINQBPPLP--AYATFTPGGAVP-----PNHLP-- 1222
 Db 1012 MSEAVQFR--TPKADSDDKMPDQALGACGKGRPLDGLSDTKPMMGSSNSPSPSTSP 1068
 Qy 1223 -----FATQRHASEYQAGLAARCAQSAACNS-- 1250
 Db 1069 LDSNMLLVIIISGIVITIVVVVLIIVFECTRTTSHQK-----KRAACKSVNGSHK 1119
 Qy 1251 ----CDALTPS-----PMOP--PPVPVPGWQYVPHNPSHMPHTSSNHOITQ 1294
 Db 1120 YGCKNDVKPPLMIHHELELKPDKSPDNVMTD--TPIFNSQDIPV----- 1169
 Qy 1295 CSSECSDSHRSQSHKQLQLEHSS-----AKQGRGHHRRAPV--VPMSENNM 1347
 Db 1170 -----DNMSDENTHQRNYSRGHESDMSLTAGRGKMPKMPOMPTQSPQOQVRYMT 1223
 Qy 1348 LAEYKQRYTSDCCNSR--EDTSCSGSGLIADAGEAP 1387
 Db 1224 STDYMPASSQCTCTDQDPGATSSSTLASSQEDSGSLP 1265

RESULT 14
 P97798
 ID P97798 PRELIMINARY; PRI; 1493 AA.
 AC P97798;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE NEOGENIN (NEOGENIN PROTEIN).
 GN NEOL.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97407661; PubMed=9264410;
 A Keeling S.L., Gad J.M., Cooper H.M.;
 T "Mouse Neogenin, a DCC-like molecule, has four splice variants and is
 RT expressed widely in the adult mouse and during embryogenesis.";
 RL Oncogene 15:691-700(1997).
 DR EMBL: Y09535; CAA70727.1; -.
 DR HSSP: P02751; ITTF.
 DR MGD: MGI:1097159; Neol.
 DR INTERPRO: IPR000531; -.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFM: PFM0041; fn3; 6.
 DR PFM: PFM0047; ig; 4.
 DR PRINTS: PR00014; FNTYPEII.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 SQ SEQUENCE 1493 AA; 163159 MW; 441DB919D5E17C0E CRC64;

Query Match 9.84; Score 731.5; DB 11; Length 1493;
 Best Local Similarity 22.64; Pred. No. 1.5e-40;
 Matches 325; Conservative 181; Mismatches 538; Indels 395; Gaps 57;

Qy 124 EQCGEGYCVAKNR-----VQGVSRHASQIQLVLRD----- 156
 Db 4 EREAGRLICTSSSRCCPPPLLLLLLLLLLPGASGAANAKSGFPRQSGASVRIPTP 63
 Qy 157 --FVPEKTRVARGETALLCGPLGKPIEPLIWIQDQVLDLAKMSGASSVRIVD 214
 Db 1014 ANGKITGTIITSTD--VNAETHD-----WYIEPVGNRLTHQIELDTPTFYIK 1062

Db 64 FYFVPEVPTVLSVSGSSVILNCS--ATSEPSNLEWKDGT-----FLNLESDDRARQLLP 116
 Qy 215 GMLLLSNV-----EPIDEGNYKCIA--QNLVTRRESSYAKLIVKVPKMKPDQVM 266
 Db 117 DGLFLLSNVYKSHKPK--DEGFCQCAVDNL--GTIVSRATKLIVAGLPRTSPQEPSPV 174
 Qy 267 LYGTATFHCSSVGGDPPPKVLKKEGNIPIVSRARILHDEK-----SLEISNITPDE 319
 Db 175 TVGNSALNCEVNNADLVFVYKMQ-----NRQQLLLDRIVKLPSTGLVSNATBDG 227
 Qy 320 GTTYCEARN--NVGQISARASLIVHAPPN-----PTKRPSN--KXVGLNVQLPCHMAG 370
 Db 228 GLYKIVSGGPPKFSDEALKVLQDEPIVDLVFLMRPSSMMKVTQGSV--LPCVVG 285
 Qy 371 NPPSPVFTKGVSTLMPFNSSHGKQYVAADGTQITDVRQEDGTYVCSAFVSVDSTV 430
 Db 286 LPAPVYKMK--NEEVLDTESSGLVLLAGGLEISDVTEDDAGTYFC-----IADGNK 338
 Qy 431 RVLQVSSYDERPPPIIQIPANQTLKSGVSATLPCRATNGSPRIKFWGHAVQACNR 490
 Db 339 TVEQAELTVNPPVPELX--QANIYAHESNDIVFECEVTKPTTVKWNKGVDPISDN 397
 Qy 491 YSLIQGSLVLDQLSDSGTYCTLSERGETNSNAATLV-----EKPGST 537
 Db 398 FKIVKHNQLVGLVKSDEBGTQCIADENVQNAQAQILILEHDAVPIPLPSTISAT 457
 Qy 538 SLRRADPSTIPAPGPTKMPNVMVSRIS--ISLRNKSQKXGAPGPIGTYVEFTSFL 594
 Db 458 TDHLA--PATIGPLSAPRDVVASIVSTRFKLWRTPASDPH--GDLNITSVFTKEGV 513
 Qy 595 QTVHVAHRQVQDVTISGLTPTSVFLVRAENTQGISVPSGLSWIKITBADFDAAS 654
 Db 514 DRERVENTSQSGEMQVITQMLPATVITFYKNAQKRG-----SGSSAPLRVE----- 562
 Qy 655 ANDLSAARTLTKGSVEL-----IDASAINASVLEMLVNSADKTYVEGLRIHTDA 708
 Db 563 -----TQPEVQLGPAPNTRATATSPISITVWTEPLSGNCE--TQNLKTYHEK 610
 Qy 709 SVPSAQTHSITVMDASAESFVGNLKKTKTYEFLFPFETIEQGPSNKTALITEDPS 768
 Db 611 GIDKED-----DIDVSSHSITLNGLKATETSEFVYVAKHNPQSTQVAVRTLSQVS 665
 Qy 769 APPNDIQIHYNTAGVWVMTPPPSGHNGNLGYKTEASNGTMVKLANMILNATTISV 828
 Db 666 AAPQLNLSLVANSKSIIVHMQPPSSITQNGQITGYKIRKASKRSQVITVGTQLSQ 725
 Qy 829 LLNMLTLGAVIVSLNPSITRAGDGPYSKPIS--LFDMPHHVHPRAHPSGTHDRHQQ 886
 Db 726 LIEGLDGTETTFNFAALVYNGGPAIDWASPTFESDLDRVPEV--PSLIL----- 777
 Qy 887 DLITHNGNIPGPDINPTTHKTTDYLSPWMLVLCVLLVLVISAISMYTKRKKHQQ 946
 Db 778 -----VFP-----LVTISLVNTPENQNVIT 798
 Qy 947 TK--ELHGLSVSDNEITALNINSKESLWIDHGRWRTADTKDGLSGESKLLSVNSQS 1004
 Db 799 VRGATGCI-----GIGSQAQTKVYKQRYTLENLDS-----SRTVTL 840
 Qy 1005 SNYNNSDG-----TDYAEVDRNLITFYNCRKSPDMPTPIATMI--IGTSS 1050
 Db 841 KAFNVVGEIGLYESAVTRPHPTDTEVLEVI-----NAPITVPDPTPMMPVGVQA 893
 Qy 1051 SETCTKTTISIS--ADKSDCTSHPSYDA-----PAGQVPAVYVKS-----NYLQVPEP 1097
 Db 894 SILSHEDTIRIWDANSPLKPKHQITDRTYTVYKNTIPANTKYKNAATLISLVYLGKPF 953
 Qy 1098 INKSE-----LPPPPHPPPSTGYAQGPS-----SR 1124
 Db 954 WLYEFSWVWYKGRSSSTWSMTAGATFELVPTSPKDVTVSEKGRPTIIVNNQPPSE 1013
 Qy 1129 KSKSAGSGI--STNQSLNASHSSSGGFSAMGVSP-----QYAVA 1169
 Db 1014 ANGKITGTIITSTD--VNAETHD-----WYIEPVGNRLTHQIELDTPTFYIK 1062

Qy 1170 CPPEYNSWPLSAVAGGTQNYQITPTNBQBPOLP--AYFATTGPGGAVP----- 1217
 Db 1063 IQARN--SSGGMPSAEVQPR--TPKADSDKMPNDQALGSAGKSLPDLGSDYKPPM 1117
 Qy 1218 -----PNHLP-----PATQRAASEYQAGLNARCA 1243
 Db 1118 SGNSPHSGPTSPLDNSMLLVIVSVGVITIVVVVYIAVCTTRTTSKQK----- 1168
 Qy 1244 QSRACNS-----CDALATPS-----PMQ-----PPVPVPEWQTPVPHNS 1281
 Db 1169 KRAACKSVNGSHKYGCKDKVPPDLWIHHERLELKPDKSPDPNPVMTD--TTPPNSQ 1226
 Qy 1282 PMHPTSNHQQICSSCSOHSRSQSHKQQLQLEBHS--SAKQGGHRRRAPVQ 1338
 Db 1227 DITPV-----DMSDSNIHQRRNSYGHESDSMTLAGRMRPKMMP 1271

RESULT 15
 000340
 ID 000340 PRELIMINARY; PRT; 1461 AA.
 C 000340:
 F 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE NDOGENIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=97312699; PubMed=9169140;
 RA Vielmetter J.J., Cheng X.N., Miskovich F., Lane R.P., Yamakawa K.,
 RA Korenberg J.R., Dreyer W.J.;
 RT "Molecular characterization of human neogenin, a DCC-related protein,
 RT and the mapping of its gene (NEOL) to chromosomal position 15q22.3-
 RT q23.1";
 RL Genomics 41:414-421 (1997).
 DR ENBL, U72391; AAC51287.1; .
 DR HSP: P02751; ITF.
 DR INTERPRO: IPR000531; .
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003005; .
 DR PFM: P00041; fn3; 6.
 DR PFM: P00047; lg; 4.
 DR PRINTS: PR0014; FNTPELII.
 DR PROSITE: P500430; TONE_DEPENDENT_REC_1; UNKNOWN_1.
 SQ SEQUENCE 1461 AA; 160015 MW; 4A0DFEECAFD2C CRC64;

Query Match 9.6%; Score 712.5; DB 4; Length 1461;
 Best Local Similarity 22.3%; Pred. No. 2.8e-39;
 Matches 313; Conservative 184; Mismatches 564; Indels 341; Gaps 48;

Qy 157 FRVPEMDTRVAGETALLEGCPKPGIPBTLWIKDGVPLDLKMGSPAGSSVRIVDGG 216
 Db 55 FLVPEVDLVSGLSGSVILNCS--AYSEPSKLEWKDGT-----FLNLVSDDRKLPDGG 107
 Qy 217 NLLISWV-----EPIDBNKYCIAQ--NLVGTSEYAKLLIVQVPMKPEKQDQMLTG 269
 Db 108 SLFTISWVSHKNNKP--DEGYTQCVATVESLCTISRAKLLIVAGLPRTSQPEPSYTAG 166
 Qy 270 QTATFSGVGGDPKPKVLMKKEBGNIPVSRARLHDEK-----SLEISNITPDEGT 322
 Db 167 NNALLNCEVNAOLVPFVMEQ-----NRQPLLDORIKLPSNGLVSNATGDOGLY 219
 Qy 323 VCEAHN--WQGIASARSLIVHAPPN-----PTKRNNKVLGVNLQVPCMASGNPPS 375
 Db 220 KCVESGSGPKPTSDVELKVLDPPEVISDLVFLQPSPLVRIQGVQLVPCVAGSLPTT 279
 Qy 375 VFYTK--EGVSTLMPFNSHGRIYAAQDTLQITDVRQDEBGTYSACFSYVDSSTYRVF 433

Db 280 LKMKKEALD-----ESSERKLVLGAGSLEISVDYDAGTYFC-----IAGQNETIE 330
 Qy 434 LQVSSVDERPPPIIIGIPANQTLPGKSVATLPCRAATNGPSRIKMFHDGHAQVGNRSI 493
 Db 331 AQAEELVQAQPEFLK--OPTNLYAHMSDNIYECVETGKPTTPVKVNGVGDWPISDYFKI 389
 Qy 494 IQGSSLVADVLQLSDSGYTCTSERGETSEWATLVKPKGTSILARAADSTPYAPPG 553
 Db 390 VKENILVQGLVKSDEGPTQCIARNDVQNAQGLIILE-----HAPATPLPSAPR 443
 Qy 554 TPXVLNWSRTSISLRWAKSQKPGAVGPIIGITYVEYFSPDLQTVIAHVRWGTQVTSI 613
 Db 444 DVVALSVSTRPKLTWRTPASDPH--GDLTYSVYTTKEGIARVENTSHPGEMVITQI 501
 Qy 614 GLTPGTSVFLVRAENTQGISVPSGLSWIKTLEADFDASANDSAAITLTKASVEL- 672
 Db 502 NLMPTATPIFYVNAQNKHC-----SGESSAPLRVE-----TPQEVQLP 539
 Qy 673 -----IDASAINASAVLEMLHVSADKXYVEGLRIHYKDAVPSAQYTHIVMDASAE 727
 Db 540 GAPNLRAYATSPYSITVITWETVPSNGSE--IQNKLYIMEKGTDKQ-----DVDVSHS 593
 Qy 728 FVYGNLKKXTIYEFLLPFPETIEQGSNKATLTIEDYSPAPPDNIQIMYQTAGWVA 787
 Db 594 YTINGLKKXTIYTSFVVAIXNHGPOVSTPDVAIVSLDVSAPRQNLSELVANSKIMIH 653
 Qy 788 WTPFPQSHGNLNLGYKIEVSAGNTMKVLAAMTLNATTSVLLNMTGAVSVRLNSPT 847
 Db 654 WQPPATQNTQYITGYKIRKASKRSVDVTEVLSQSLQLEGLDGRGTETYNFVAALT 713
 Qy 848 KAGDGYSPKPS--LPMDFTHVHPRAHPSGTGDRHGQQOLYHNWNGIIPGSDINTPT 905
 Db 714 INTGPTATWLSAETPESDLDETVPVEY--PSLLH-----VRF-- 749
 Qy 906 HKKTDYLVSGWMLVLCVILVLISAAISVYFKRQMKV--ELHLSVYSDWEITA 963
 Db 750 -----LVTSVWSTPPEQNLVYVGTATG-----G 776
 Qy 964 LWNKESMIDHHHGNRTADTDCGLSGSKLLSHVNSQSNVNSDGS----- 1013
 Db 777 GSGPHAQIKVDYQRYTILENLPDS-----SHVYTLAFLNNGEGLPVSATVR 828
 Qy 1014 -----TDYAVDTANLTIFYNCKSPDNPPTIATMI--IGTSSETCTATTSIS--ADKSGT 1068
 Db 829 PHPTDSEVQLVVI-----NAPYTPVDPPTPMMPFVGQASILSDITRITWADNSLPK 881
 Qy 1069 HSPSTD-----FAGQVPVAVVKS-----NLYQVPEVINSEF----- 1103
 Db 882 HKKTDGRTYVKKTNIPANTKYKNAATLSLVTLGLKPLNTLYFSVMYTKGRSSSTV 941
 Qy 1104 -----LPPPPHPPSSSTYGAQGSPE-----SRKSSKAGSGI--STNQSL 1145
 Db 942 SMATAGTTELPTVSPKDVTVYSKBEKPTILVNWQPPSDANGKITGYITSTD--V 998
 Qy 1146 NASIH-----SSSSGSGSANGVSPQATV--ACPPEN 1174
 Db 999 NATHDHWIETPVGNRLTHQIQLDTPYFKIQARNSGKMGPMSEAVQFPTKADSSD 1058
 Qy 1175 VYNSPLASAVAGTQNYQITPTNBQBPOLPAYPATTPGGAVPVPHL----- 1221
 Db 1059 KMPNDQASGSGKSLRDLGSDGPKPPMGSNSPHGSPPTSPLDNLNLIIVSVGVITIV 1118
 Qy 1222 -----PFATQRAASEYQAGLNARCAQSRACNSCA-----LATPSMQP----- 1263
 Db 1119 VVILIAVCTRTTSHQK-----KRAACKSVNGSHKYGCKDKVPPDLWIHHER 1169
 Qy 1264 -----PPVPVPEWQTPVPHNSPHMPTSNHQQICSSCSOHSRSQSHKQQLQ 1315
 Db 1170 LELKPIDKSPDPNIMTDTPIRNSQDITPV-----DMSDSNIHQRRNSY 1215
 Qy 1316 EHHGSS-----AQKRGHRRRAP-----VQPCMESENENMLAEYQRTYDCCNSRS 1365

Db 1216 RGHSEDSMTLAGRGRMRPKMMMPFDSQPPQPVISARPIBSLDNPRHHFHSLSLASPAR 1275

Oy 1366 EGDTCSCSGSCLYARAGEPAP 1387

Db 1276 -----SHLY-HPGSPWP 1286

Search completed: January 22, 2001, 12:51:13
Job time: 1874 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:17:03 ; Search time 233.01 Seconds
(without alignments)
202.659 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTPVKNP.....RSLLSNWSGGTSSQPAHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq.36:
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	7272	100.0	1381	20	Y13564	Drosophila Robo 2
2	7256.5	99.8	1380	20	Y08402	Drosophila sp. ROB
3	1786.5	24.6	1395	20	Y13563	Drosophila Robo 1
4	1786.5	24.6	1395	20	Y08401	Drosophila sp. ROB
5	1498	20.6	1651	20	Y13566	Human Robo 1 polyp
6	1491	20.5	1649	20	Y08404	Human ROB1 protein
7	1344.5	18.5	1297	20	Y13565	C. elegans Robo po
8	1344.5	18.5	1297	20	Y08403	C. elegans ROBO pr
9	1261	17.3	753	20	W83927	Human T85 protein.
10	557.5	7.7	1728	12	R13144	Deleted in Colorec
11	551.5	7.6	1447	16	R69553	Deleted in colorec
12	551.5	7.6	1447	20	Y33498	Human DCC protein.

13	548.5	7.5	1192	19	W57900	Protein of clone C
14	543.5	7.5	1028	19	W29667	Homo sapiens DL165
15	538	7.4	1571	19	W42087	Human Down syndrom
16	538	7.4	1910	19	W42086	Human Down syndrom
17	536.5	7.4	1018	15	R63759	Human contactin (E
18	536.5	7.4	1018	17	R87028	Human contactin.
19	531.5	7.3	1496	20	W81030	Melanoma associat
20	531.5	7.3	1496	21	Y70469	Human p53 target s
21	524	7.2	1018	18	W06485	Rat contactin liga
22	509.5	7.0	1299	21	Y40439	Human Nr-CAM prote
23	509	7.0	1257	20	W74152	Human LI cell adic
24	495.5	6.8	1897	21	Y81785	Human protein tyro
25	495.5	6.8	1897	21	Y56100	LAR tyrosine phosph
26	495	6.8	1304	19	W59994	Human neural cell
27	493.5	6.8	1911	16	R71726	Human PTP-OB. Hom
28	493.5	6.8	1911	18	W27225	Human protein tyro
29	493.5	6.8	1911	20	W94027	Human protein tyro
30	484	6.7	3117	21	Y53667	Sequence gi/332818
31	475	6.5	1242	19	W52287	Rattus norvegicus
32	473.5	6.5	434	20	Y13567	Human Robo 2 polyp
33	473.5	6.5	434	20	Y08405	Human partial ROBO
34	448.5	6.2	4412	21	Y53666	Sequence gi/101742
35	438	6.0	1225	19	W52289	Homo sapiens cdo t
36	431	5.9	1501	16	R72858	Rat receptor type-
37	420.5	5.8	761	17	R92255	Neural cell adhesi
38	415.5	5.7	1251	19	W37778	Rattus norvegicus
39	412.5	5.7	1070	18	W08747	Human colon carcin
40	388.5	5.3	848	21	Y88565	Human NCAM 140kd i
41	382	5.3	1853	21	Y53668	Protein 608 sequen
42	382	5.3	2387	21	Y53665	Mechanical stress
43	382	5.3	2597	21	Y53664	Mechanical stress
44	370.5	5.1	1125	19	W52288	Rattus norvegicus
45	370.5	5.1	1139	19	W37779	Rattus norvegicus

ALIGNMENTS

RESULT 1

Y13564
ID Y13564 standard; Protein; 1381 AA.
XX
AC Y13564;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W09925833-AL.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 38WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RBGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-338003/28.
DR N-PSDB: X55768.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure: Page 34-38; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.

XX
 SQ Sequence 1381 AA;

Query Match 100.0%; Score 7272; DB 20; Length 1381;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GENPRLIEHPMDTVKNDPFFNCQAGNPPTPIQWFGREGLKDTGSHRIMLAGGL 50
 DB 1 genpRLiehpmdTVkNDpffncQagNPpTPIqWfGREGLkDTgSHrImLaGGL 50
 61 FFVFKVHSRRSDAGTYCAKNEFGVARSNATLQVAVLREDFLEPANTYVQAGEVAL 120
 DB 61 fFvFKvHsRRsDagTyCAkNEfGVaRSnATlQvAVlREdFLEpANTyVQAGeVAL 120
 121 MECGAPGSPGPQISWRKNGQTLNVLNKRIRVDGGLAQEARGSDGTCVVMVW 180
 DB 121 meCGaPgsPgpQISwRKngQTLnVlNKRIRvDgGLAQeARgSDgTCvVMVw 180
 181 GTRESATFLKVRPFLIRGPQNVAVGSSVFCRIGGDPIDVLRWRTASGNGWL 240
 DB 181 gTReSaTfLkVRpFlIRgpQnVAVgSSvFcRiGGdPiDvLRwRtASgNGwL 240
 241 RFSWSLHSSAGSRVHLEDRLKLDVLTDMGSEYTCEDANVGGITATGLTWHAPKFV 300
 DB 241 rFSwSLHssAGsRVhLEdRLkLDvLTdMGSEyTCeDANvgGiTaTGLtWhApKfV 300
 301 IRPKNQLVEIGDEVLFECQANGHPRTLWSVEGNSLLPGYRDMGEMTLTPBGRSVL 360
 DB 301 iRpkNqlVeiGDevLfEcQanghPrTLwSvEGnSllPGyRdMGEMtLTPBgRsvL 360
 361 SIARFARSDGKVTNCALNAVGSVSRVTVSDTQFELPPPILEGQPNQTLPVKSIIV 420
 DB 361 siARfARsDgKvTncAlNAVgSVsrVTvSDtQfELppPIlEGqPNqTLpVksIiv 420
 421 LPCRTLTGPVQSWYLDGPIVDQEHERRNLSDAGALITSDLRHEDGLEYTCVASRN 480
 DB 421 lPcRtLTgPvQswYlDgPiVdQehErRnLsDagAlITsDLRHedGleyTcVaSRn 480
 481 GKSNSYSLRLDTPTNPNKIFRAPELTSYPPGPKPMQVEKGENSVTLWSRNMKGVS 540
 DB 481 gKSnsYsLrLdTpTnPNkIFrApELtSYppGpKPMqVEkGEnSVtLwSRnMKGvS 540
 541 SLVGYVIEFMGKNETDGNVAGTRVQNTFTQTLGLGVNYYFLIRAESHGSLSPMS 600
 DB 541 slVgyVieFMgKneTDgNvAGtrVqNtFTqTLgLVnYYfLiRaEShGslSPms 600
 601 EPIVTGTRYFNSGLDSEARSLSGDVLSNASVDSMTKLTWLIINGKYVEGFYV 660
 DB 601 ePiVtGtryfNSglDseARslSGdVLSnasvDsmTKLTwLIiNGkyVeGFyV 660
 661 ARQLPWPVNPVNTNPNLLSGTSTASASASASALISTKPNIAAGKRDEGTNOSG 720
 DB 661 arQLpWPvNPvNTnPNlLlSGtSTaSaSaSaSaLiStKpNIaAGKRdEGtNoSG 720
 721 GGAPTPLNKTMLTILNGGASSCTITGLVQYTLFFVYPTPKVSEKSPNSRIARTL 780
 DB 721 gGaPtPlNkTmLTiLNgGassCtITglVqYtLFFvYpTPkVSEkSPnsRIarTL 780
 781 EDVPEAPTEAGMEALLNSAVFLKWKAPELKDRHGVLNTHVYRGIDTAHNFSLIN 840
 DB 781 eDvPeApTEaGMEaLLNsAvFLkWKaPElKDRHgVLNthVYrGiDTaHnFslIN 840

QY 841 TIDASPTLVLANLLEGVMTYVGAAGNAGVGPYCPATLRLDPTKRLDPIQRDHF 900
 DB 841 tIdasptLVlanLlEGvMTyVgAAGnAGvGPYcPATlRLdPTkRLdPIQRdHF 900
 901 NDVLTPQWPIFILLGALLVLMISFGAMVFKRHMMKQASLNTNRGHTSDVLKMPSL 960
 DB 901 nDvLTpQwPIfILLgALLvLMISfgAMvFkrHmMkQaSLnTNRghTSDvLKmpSL 960
 961 ARNGNWFYKSDTGKVMWRPSPGGSLEMQKHADIAPVCPAGSPAGGTSSTSGSGGA 1020
 DB 961 aRNgNwFYkSDtGkVMwRPSPgGSLEmqKHAdIaPvCPaGSPaGGTSstSGSGga 1020
 1021 GSGASGGDILHGGHSGERNQQRVGEYSNIPTDAEVSFGKAPSEYGRHGNASPAYAT 1080
 DB 1021 gSgASgGDilHggHSGeRNqQRvGEysNIpTDAeVSfGkAPsEYGRhGNasPayAT 1080
 1081 SSLSPHQOQQQQPPTVQORPVGVLGQPMHPHYQQOQQOQQOQQOQQOQQOQQOQQ 1140
 DB 1081 ssLSpHQoQQQPPtVqORpVGvLGQpMhPhYqQOQQOQQOQQOQQOQQOQQOQQ 1140
 1141 LPPSNLYQMTSTSEIYPTNTPGSVSVSEGYITPDKQRHIIHTEMLKNSCHTYEAPG 1200
 DB 1141 lPpsNlyQmTstSeIyPTnTPGSvSVSEgyITpDKQRhIIHteMLKnsCHtyEaPg 1200
 1201 AKQSPSTISQPSAVRRQQLPNCISGRSAFVKVLTQDGKNQMLDLSGSMCTWGLA 1260
 DB 1201 aKqSPstISqPSaVRRqQLpNCiSgrSAfVKvLTQdgKNqMLdLSgsmCTwGLa 1260
 1261 DSGCGSPSPHMLMSHEDAHLYHTADGDLDMERLYVVKDEQPPQQOQLTFLVPQH 1320
 DB 1261 dSGcGSPspHmLMSheDAhLyHTadGdLdMerLyvVKdeQppQQOqlTfLvpQH 1320
 1321 PASHGLQSMRWSTSRSSKNGQCEKPESELYIAPGSVASERLSLNSGSGTSSQAGHN 1380
 DB 1321 pAShGLqSMrWstSRSSKngQceKpEsELyIaPgSVaSErLSlNSgSGtSSqAGhn 1380
 QY 1381 v 1381
 DB 1381 v 1381

RESULT 2
 Y08402
 ID Y08402 standard; Protein; 1380 AA.
 XX
 AC Y08402;
 XX
 DT 24-JUL-1999 (first entry)
 XX
 DE Drosophila sp. ROBO2 extracellular domain protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 KW cell morphology; screening assay.
 XX
 OS Drosophila sp.
 XX
 PN W09920764-AL
 XX
 PD 29-APR-1999.
 XX
 PF 20-OCT-1998; '98WO-0522164.
 XX
 PR 14-NOV-1997; '97US-0971172.
 PR 20-OCT-1997; '97US-0062921.
 XX
 PA (REBC) UNIV CALIFORNIA.
 XX
 PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
 XX
 WP1; 1999-312615/26.
 DR N-PSDB; X57251.
 XX
 PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 52-56; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1380 AA;

Query Match 99.8%; Score 7256.5; DB 20; Length 1380;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1380; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 GENPIREHMPOTTVKNDPTFNCQAEENPTPIQWFKDRELKTDGSHRMLPAGGL 50
Db 1 genpirehmpotvknndptfncqaeenptpiqwkdkrelktdgshrlmpagl 50
Qy 61 FFLVLIHSRESADAGTYWCAENEFVARSNATLQVAVRDEFLEPANTVACGEVAL 120
Db 61 flvlvlihsresadagtywcaenefvarsnatlqvavirdeflepantrvacgeval 120
Qy 121 MECGAPRSGPEPISWKNQNTLNLVGNKRIRVDGNSLQAEQSGDGRYCVKNVW 180
Db 121 mecgaprgsgepisiwknngntlnlvgnkrrivdgngslqaeqsgdgrycvknv 180
Qy 181 GTRESATFLKWHVRFILRGPQNTAVGSSVFCRIGGDLPLVWRRTASGGNPL 240
Db 181 gtresatflkwhvrflirgpnqntavgssvfcrriggdlplvwrtrtasggnpl 240
Qy 241 RKFSWLHSSGVRVHEDRSKLDDVTLEDMEYTCEDANVGTITAGLTITVAPKVF 300
Db 241 rkfswlhssgvrhvhdslkddvtledmeytceadanvgtitagltitvapkfv 300
Qy 301 IRPNKLVEIGDEVLFECQANGHPRLPYLNSVEGNSLLPCTYDGRMEVTLPEGSVL 360
Db 301 irpnklveigdevlfecqanghrpilylnsvegnssllpctydrmevtlpegsvl 360
Qy 361 SIARFARSDSGKVTCTNALNAVGSVSRVSVDTQFELPPPIRQGPVNTLWKSIV 420
Db 361 siarfarsdsgkvtctnalnavgssvrsvsvdtqfelpppirlqgpvntlwksiv 420
Qy 421 LPCRITLTPVPVNSWLDGIPIDVQEHERRNSDAGALISDGLHDEDEGLITCVASNR 480
Db 421 lpcritltpvpvnswlldgipidvqeherrnsdagalisdglhdedeglytvcasnr 480
Qy 481 GKSSNSYLRIDTNPNIKFRAPELSTYPPGPKQPMQVEGNSVLSWTRSNKVGSS 540
Db 481 gkssnsylridtntpnkifrapelstypgpkpmqvegnsvlswtrsnkvgs 540
Qy 541 SLVGYIEMFNKNETDGVAVGTRVQNTFTTQGLLPGVNYFFLIARENSHGLSLPSPS 600
Db 541 slvgymfnknetdgvavgtvqntfttqglpgvnyffliarensghslslpsps 600
Qy 601 EPIVTGTRFYNGLSGLSEARSLISGDVVELSNASVDSSTSMKLTWQINGKIVVEGYTY 660
Db 601 epivtgtrfynsglslsearslisdvvelsnasvdsstsmkltwqingkvvegyty 660
Qy 661 ARQLPNIPIVNNPAPVTSNPLLGSTSTASASASALISTKPNIAAGKRGDETQSG 720
Db 661 arqlpniipivnnpavpvtstnpllgststasasasalistkpniaagkrgetqsg 720
Qy 721 GYAPTPLNTKYRMLTILNGGASSCTITGLVQTYLFFVYFPTKSVGKPSRSRIARTL 780
Db 721 gyaptplntkyrmltilnggassctitglvqtylyffvyfptksvgkpsrsriartl 780
Qy 781 EDVPEAPYMEALLNSAVFLWKAPELDRHGVLNHYVIRGIDTAHNSRILTW 840
Db 781 edvpeapyymeallnsavflwkapeldrhgvlhyvirgidtahnsriltw 840

Qy 841 TIDASPTLVLANLTGVMYTVGVAAGNAGVGPYCPATLRLDPTIKRLDFFINRDHV 900
Db 841 tidasptlvlanltgvmtyvgvaagnagvgycpatrlrldptikrldffingrdhv 900
Qy 901 NDVLTPWPFVILLGAILAVLMSFGAMVFKRHHMMKQSAINTNRGNTSDVLKMPELS 960
Db 901 ndvltppwfvillgailavlmisfgamvfkrrhmmkqsaintrngntsdvlkmpels 960
Qy 961 ARNGNGTVLWDSSTGMMWRPSPGGSDLEMQKHADIAPVPCGAPSGAGGTSQSGSGA 1020
Db 961 arngngtvldsstgmmwrpspgsdlemqkhiadiapvpcgapsgaggtssqsgga 1020
Qy 1021 GSGASGGDDIHGGHSGERNQRYVGEYENIPTDYAEVSFGKAPSEYGRHGNAPATYAT 1080
Db 1021 gsgasggddihgghsgernrqryvgeyeniptdyaevsfgkapseygrhgnaspatyat 1080
Qy 1081 SSILSPHQQOQQQQQPRYQORPVPGYGLORPMHPHYQQOQQOQQOQQOQQOQQOQQO 1140
Db 1081 ssilspqhqqqqqqqpryqorpvpgyglorpmhphyyqqqqqqqqqqqqqqqqqqqq 1140
Qy 1141 LPPSGLYQMSVTSELYPTNPGSPSVSEYQYTYPKDQRIHITENKLSNCHTYEAPG 1200
Db 1141 lppslyqmsvtseilyptnpgspsvseyqytypkdqrihi-enklsnchtyeapg 1199
Qy 1201 AKQSSPISQFASVTRQQLPPWCISGRESARFVLTDTQCKNQNLNLLDSSSMCYGLA 1260
Db 1201 akqsspissqfsvtrqqlppncisgresarfvltdtqckngqnllddssmcygla 1259
Qy 1261 DSGCGSPSPMAMLSHEDENALHYLTADGDDLMERLYKVKDQPPQQOQQQLIPVQPH 1320
Db 1261 dsgcgspspamlshedenalhyltadgdldmerlykvkdeppqqqqqlipvpqh 1319
Qy 1321 PACHGLQSWRQSTSSSKRWQCEKESLIIAPGSVASERSLSNSGSGTSSQPGHN 1380
Db 1321 pachglqswrqstssrknqceklpsellyapgsvaserslsnsgsgtssqpghn 1379
Qy 1381 V 1381
Db 1380 V 1380

RESULT 3
Y13563
ID Y13563 standard; Protein; 1395 AA.
XX
AC Y13563;
XX
DT 30-JUL-1999 (first entry)
XX
DE *Drosophila* Robo; polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
XX
OS *Drosophila* sp.
XX
PN W09925833-AL.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-0524327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RBC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55767.
XX
PT Modulation of Robo-Comm polypeptide interactions

XX

PA (RBCG) UNIV CALIFORNIA.

XX

PI

Goodman CS, Kidd T, Mitchell KJ, Tear G;

XX

DR WPI; 1999-312615/26.

XX

DR N-PSDB; X57250.

XX

PT Robo polypeptides, a new immunoglobulin superfamily member

XX

XX

PS Claim 1; Page 45-49; 80pp; English.

XX

CC This invention describes novel Robo (roundabout) polypeptides, involved

CC in nerve guidance which have been isolated from *Drosophila* sp.,CC *C. elegans*, human and murine samples. The products of the invention can

CC be used to raise anti-Robo antibodies, which can be used to modulate cell

CC function or morphology. The Robo polynucleotides and fragments are useful

CC as probes and primers and for production of the Robo polypeptides. The

CC probes and primers are also useful in screening assays.

XX

Sequence 1395 AA;

Query Match 24.6%; Score 1786.5; DB 20; Length 1395;
Best Local Similarity 30.1%; Pred. No. 1.le-101;
Matches 441; Conservative 242; Mismatches 507; Indels 275; Gaps 39;

Qy 2 ENPRIIEHMDTVPKNDPPTFNCQAEQNTPTIQWFKGRLEKT-DTGSRRMLPAGGL 60

Db 54 SPQRHIEPTDLVKKNEPATLNCKVEKPEPTLEWFKDGEPTVNEKKHVRQFKDAL 113

Qy 61 FFLAVLHRSRESADAGTYCEAKNEPGVARSNRATLQAVLRDEFERLEPANTRVAGQVAL 120

Db 114 FFYRTNQPKGKEGSEYCWCAKRVGSGVASHALQIARLDRDFVEPKTRAVKAGTAL 173

Qy 121 MEGCAGPGRSGEPPIQSHKNG-----QTLNLWGKSRIRIVDSQGNALQARQSDQGYQC 174

Db 174 LECQPKGIPEPTLIWIKDGPIDDLKAMSFGASSRVIVDGNLISVPEIDEGNYKC 233

Qy 175 VYVNVTVRESATAFKLVNRPFLRGPQNTQAVGVSSVFCRIGDGLDPLVLRERTAS 234

Db 234 IAGLPTTRESSYAKLVQVQKPIFKPEKQDQMLVYGTATFCHSGVDPKPKWK--e 291

Qy 235 GGNMPLKSKMLASAGRVHLEDRLALDVTLEDGMYTCADNAVGGITATGILTVH 294

Db 292 EGNIPVSRARILH-----DEKLEISNITDGTGTYCEAHNNGVSIASRALVH 342

Qy 295 APPKFIKPNKQNLVEIDDEVLFECQANGHPRTPLVSVBNGSNLLPGTRDGMVTLTP 354

Db 343 APPKFIKPNKQNLVGLVQPCMASGNPPSVFVKEGYSTLNFPOSSHGRYVA--- 399

Qy 355 BGKSVLSIARFARDSGKVVTCNALNVAGSVSSRTVYSVDTPQLPPIIBQGVYQTL 414

Db 400 -ADTLQIGTDVQDEGYV-CSAFSVSDSTVRFVIGVSSVDPPILQIQAQNTLP 457

Qy 415 VKSVIVLPCRLTGPVQVSNYLDGIPIDVQHERRNLSDAGALTISQLRHDEGLTYC 474

Db 458 GKSATVLCRAQNGSPRIKVFHDGHA--VQAGRYSLIGSSLRVDQLG-ADSTYTC 514

Qy 475 VASNRNKKSSNGYLRDLTPTPNWIKFPAPELSTPGPKQKPMVKGNSVTLSTWRS 534

Db 515 TASRGYETSVAATLVKPGSTSL--HRAADSTYPPAGPTKPLNVSTSLSLRWKS 572

Qy 535 NKVGS--SLVGYIEMFGKNEGDGWAAGVIRQNTFTPTGLPGVNYFFLIRAGNSHG 592

Db 573 GKQPGVPPPIGYTVEYSPDLGTGVVAHRVYDGTQVTSIGLTGTSYVIRAEANTG 632

Qy 593 LSLSPSPMSEPIVGTFRFN--SGLDLSARASLSGDVLSNASVSDTSKMLTWQI-- 648

Db 633 ISVSPGSLSVIKTLEADFAASANDLSAAR-TLITGKSVELDASALNASAVLEMLHV 691

Qy 649 -INGKIVEGTYTARQLPNIIVNPAPVTSNTWPLLGSTTSAGASASASALSTKPMIA 707

Db 692 Sadekyegrlrhykdasvp----- 711

Qy 708 AAGNRQGETNDSGGGAPPTLNTKYRMLTLNGGASSCTITGLVQTLTYEFFIVPYKSV 767

Db 712 -----SAGYHSLVMD-ASAEFVGNLKKYKYEFFITPTFETI 750

Qy 768 EKGPSNRRIARTLEDVPSAPYQGEALLNSSAVFLKNAPELKRHGVLLNHYVIRGI 827

Db 751 EGGSANKALTIEDVPSAPPDNIQIMYQGTAGVVRVTPPSQHNGALYSYKLEVV-- 807

Qy 828 DTANHSRILNLTIDAASPTILVLANITGEMVTYVGAAGNAGVGPTCPATLRLDPT 887

Db 808 -SEGNTKVLAMTLNATTSVILNLTGAVYVSLNFTKAGDGPYSKPLSMDP-T 865

Qy 888 KRLDP-----FINQRDH--VNDVLTPQWFLILGAL 916

Db 866 HHVHPPRAHPSGTHDGHGQDITYNHANGIPDGPINPTTHKKTDLGSPVNLVCI 925

Qy 917 LAVMLSPG-AMVTVKRRKRRQKQ-SALNTMRGNHNTSDVLRKPSLRNNGNFWLSD 974

Db 926 LLVLSAASIAISVYFKRHGKTELGHSVSDN-----ETALNLSKSLVLDHHR 979

Qy 975 GHWVRPSGGDSLEMQKHADYAPVPCGAPSGAGGTSSGGSGGASGGDIGHGG 1034

Db 980 -----WRTAETDKD-----SGLSKSLSHVW 1001

Qy 1035 GSRERQQRVYGEYSNIP--TDYAEVSSFGKAPSEYGRHGNASAPYATSSILSPHQ 1092

Db 1002 SEGSN-----YNSDGTGYAEVDTRNLTTYNCRKSPDNPCTYATTMLGTSSSET 1054

Qy 1093 QQRVYQARVPTGTLQRPMHMYQQQQHQQQQQHQHQH--QALQHQQLPQVYQW 1150

Db 1055 TKTSLSADKDS-THSPSYSDAFAGYPAVPPVSKSYLGYPEVINSEFLP----- 1106

Qy 1151 STSEIYPTNPFGPSRSVSYDQYIPKDKRHHH-----ITENKLSNCHTYEAPAKQS 1204

Db 1107 -----PPEHPPSTSYAGSGPSSRSKSSAGSGISTGSLNASHSSSSGSYFE 1159

Qy 1205 SPSSSQFASVRQQLPPP-----NCISIGRESARFVNLTDQKNQNLQDLSGMSY 1256

Db 1160 WYSPGYAVA-----CPENNVSNPLSAVAGTGTYGLTPTNQHPPQLPAY----- 1206

Qy 1257 NGLADSGGGSPHMLMSHDEHALYHTA-----DGLDLMERLTVKVDQPP 1307

Db 1207 --FATGPGGAVPNHPIFATGRHAASEYQAGLNAARCAAGRACNSCDALTPSPMP 1264

Qy 1308 QQQQQLPLV-----PQPAEHLQSWNRQTSRSRNGQCIEKPESLIYAP 1355

Db 1265 P-----VPPEYVYQVPHNSHPMTSSNHQY-----GQSSCED--HSR 1304

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

Qy 1356 GSVAESRLLNSGTSQSPAGHN 1380

Db 1305 SSGCHKRLQLIEHGSSEAKRGHH 1329

PF 13-NOV-1998; 98NO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-338008/28.
DR N-PSDB: X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure: Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commisureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1651 AA;

Query Match 20.6%; Score 1498; DB 20; Length 1651;
Best Local Similarity 28.9%; Pred. No. 7.8e-64;
Matches 410; Conservative 221; Mismatches 526; Indels 260; Gaps 42;

QY 4 PRIIEHMDTIVKADPTFNCQAGNPTPIQWFKDRELKTDTC---SHRMLPAGUL 60
DB 68 PIIIEHMDTIVKADPTFNCQAGNPTPIQWFKDRELKTDTC---SHRMLPAGUL 127
QY 61 PFLVWVHRSR-ESDAGTYWCEANKEFVARSNATQLVAVLRDPEFLPANTVWAGVEA 119
DB 128 PFLVWVHRSR-ESDAGTYWCEANKEFVARSNATQLVAVLRDPEFLPANTVWAGVEA 187
QY 120 LMECGAPRSGPEPISWRKNGQTLMLVGNKRIIR-VGGWLAIEARQSDGRTQVCKN 178
DB 188 LMECGAPRSGPEPISWRKNGQTLMLVGNKRIIR-VGGWLAIEARQSDGRTQVCKN 244
QY 179 VGTRESATAPLVKWHVRFLIRGPNQITAVGSSVVPQCRIGDPLFVLWHRITASGNN 238
DB 245 VGTRESATAPLVKWHVRFLIRGPNQITAVGSSVVPQCRIGDPLFVLWHRITASGNN 302
QY 239 PLAKFSWLSASGRVHYLDRSLKLDVLEDEGYTCEDNNAVGGITATGLTVHAPPK 298
DB 303 PLAKFSWLSASGRVHYLDRSLKLDVLEDEGYTCEDNNAVGGITATGLTVHAPPK 352
QY 299 FVIRPKQMLVIGEDVLEFQAGNRPRTPLVYSEGNSSLL---PYGRDMEVTLTPE 355
DB 353 FVIRPKQMLVIGEDVLEFQAGNRPRTPLVYSEGNSSLL---PYGRDMEVTLTPE 412
QY 356 GRSVLISARFARSDGSKVTVNALNAVGSVSSRTVSV-DTQELPPLPPIBQGVNQLP 414
DB 413 GRSVLISARFARSDGSKVTVNALNAVGSVSSRTVSV-DTQELPPLPPIBQGVNQLP 467
QY 415 VKSVIPLVCLRTGLTPVPQSWYLDGIPIDQVHERNLSDAGALTISDLQRHEDBLTTC 474
DB 468 VKSVIPLVCLRTGLTPVPQSWYLDGIPIDQVHERNLSDAGALTISDLQRHEDBLTTC 525
QY 475 VASNRNKGSSWGSYGLRD-----TPTNPKFPRAPLSTYPPGPKQPMKVEKEN 525
DB 526 VASNRNKGSSWGSYGLRD-----TPTNPKFPRAPLSTYPPGPKQPMKVEKEN 575
QY 526 SVTLSSWRNKSNGSSSLGTVIEMPGKNETGWDVAVGTVRQNTITPTGLGLPGVNYFLFI 585
DB 576 SVTLSSWRNKSNGSSSLGTVIEMPGKNETGWDVAVGTVRQNTITPTGLGLPGVNYFLFI 634

QY 586 RAENSRLSLSPWSEPI-TVOTRYFNGLDLSEARASLLSGDVNLSASVSDTSNKL 644
DB 635 RAENSRLSLSPWSEPI-TVOTRYFNGLDLSEARASLLSGDVNLSASVSDTSNKL 693
QY 645 TWQI-INGKYVGPVTYARQLPNPFINNPAPVTSNTNPLGSSSTASASASALISTK 703
DB 694 TWQI-INGKYVGPVTYARQLPNPFINNPAPVTSNTNPLGSSSTASASASALISTK 712
QY 704 PNIAAGKRGSTWQSGGGAPTPLNKTMLTLNNGGASCTIGLVTLYTEFIVVF 753
DB 713 PNIAAGKRGSTWQSGGGAPTPLNKTMLTLNNGGASCTIGLVTLYTEFIVVF 754
QY 764 YKSVGKPSNRIARTLEDVSEAPYGEALLI---NSSAVFLWKAPELDRHGLNTH 821
DB 755 YKSVGKPSNRIARTLEDVSEAPYGEALLI---NSSAVFLWKAPELDRHGLNTH 814
QY 822 VIVGIDRAHNFSLINVTIDAASPTLVIANLTGKVTYVGVAGNAGVGPVYATL 881
DB 815 VIVGIDRAHNFSLINVTIDAASPTLVIANLTGKVTYVGVAGNAGVGPVYATL 869
QY 882 RLDPITKRLD---FINQRDHNVLQTPWFILLAGALVLMSPGMVFKKHHQWQ 939
DB 870 RLDPITKRLD---FINQRDHNVLQTPWFILLAGALVLMSPGMVFKKHHQWQ 926
QY 940 SALNTNHNHTSDVLMKPSLS-----ARNNGVYLDSSDGMWNPSPGGSDLEMKQ 992
DB 927 SALNTNHNHTSDVLMKPSLS-----ARNNGVYLDSSDGMWNPSPGGSDLEMKQ 974
QY 993 HIADYAPVCGGAGSG---ASGDDHGHGHSERNQRTV---1044
DB 975 HIADYAPVCGGAGSG---ASGDDHGHGHSERNQRTV---1034
QY 1045 ---GEYSNIPDTAEVSSFGKAPSEYGHGAS---PAPATYSSLSFPHQOQQQQPRTQ 1098
DB 1035 ---GEYSNIPDTAEVSSFGKAPSEYGHGAS---PAPATYSSLSFPHQOQQQQPRTQ 1089
QY 1099 QRPVPTGLGAPM---PHYQQOQH---QQOQAQTHQOALQOHLPPSNITQMSY 1153
DB 1090 QRPVPTGLGAPM---PHYQQOQH---QQOQAQTHQOALQOHLPPSNITQMSY 1141
QY 1154 SEIYPTNTPGSRVSEYQYTFYKDKRHHITENKLSNCHTYEAPAGKSPSSISQAS 1213
DB 1142 SEIYPTNTPGSRVSEYQYTFYKDKRHHITENKLSNCHTYEAPAGKSPSSISQAS 1164
QY 1214 YRQQLPFPNIGRESARFVNLTDGKNQNLDDGSSMCTNGLADSGGSGSPHMAN 1273
DB 1165 YRQQLPFPNIGRESARFVNLTDGKNQNLDDGSSMCTNGLADSGGSGSPHMAN 1192
QY 1274 LMSHEDERHAIHTADGLD-----MERLYVKKVQDQPPQOQQQLPLVQVPAEHL 1326
DB 1193 LMSHEDERHAIHTADGLD-----MERLYVKKVQDQPPQOQQQLPLVQVPAEHL 1252
QY 1327 QSWNRQSTRSRKNGQ-----CIKPESELITAP 1355
DB 1253 QSWNRQSTRSRKNGQ-----CIKPESELITAP 1289

RESULT 6
ID Y08404 standard; Protein; 1649 AA.
XX
AC Y08404;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
PN W09920764-AL.

PN W09925833-A1.
 XX
 PD 27-MAY-1999.
 XX
 PF 13-NOV-1998; 98W0-US24327.
 XX
 PR 14-NOV-1997; 97US-006543.
 XX
 XX (GREG) UNIV CALIFORNIA.
 PA
 PI Goodman C, Kid I, Mitchell KJ, Russell C, Tear G;
 XX
 DR WPI; 1999-338008/28.
 DR N-PSDB; X55769.
 XX
 PT Modulation of Robo-Comm polypeptide interactions
 XX
 PS Disclosure; Page 38-39; 56pp; English.
 XX
 CC The invention relates to a method for modulating the amount of Comm
 CC (commisureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.
 XX
 SQ Sequence 1297 AA;

Query Match 18.5%; Score 1344.5; DB 20; Length 1297;
 Best Local Similarity 25.8%; Pred. No. 1.6e-74;
 Matches 386; Conservative 216; Mismatches 486; Indels 353; Gaps 46;

QY 74 PRIIEHFMOTIVFANDPFFNCQABGNPTPIQWKGDEL---KTDGSHRMLPAGGL 60
 DB 30 LVIEHPIDVSVSGPATINCQKPS-takitvkgdgovntkeqnsnvlvdtgsl 88
 QY 61 FFLKVIHRS--RESGATGYWCANKEFGVARSKNATLVQVNLVEFLEPANTRVAGQEV 118
 DB 89 filkvsngkngkdsdagayvcasnehgveksnegsklamlredfrvprtvqlgqem 148
 QY 119 ALMDGAPGSPPEQISWRKNGQTLNLVGNKRIRIVDGGNLIAEQARQSDGRTQVCVN 178
 DB 149 avlecsprgfppepvsvrkdokelrigdmprylthsgdglidvdvrsdgtvgcann 208
 QY 179 VVGTRASATFLKVRHVFLLIRGPNQATVAVGSSVFGCGRIDGDLVNLKRTASGGNM 238
 DB 209 mvgersvsnparlsfepkfeqekdmtvdgavldfordvtdgdpqgtkw--knepm 266
 QY 239 PLAKFVSLHLSAGRVHYLED-RLSKLDVTDLEMGETTCADNAVGGITANGILVHAPP 297
 DB 267 prt-----raylaktngrlrlervqsgdegyvcyarnpagtleasahlvraqpp 316
 QY 298 KFPVIRKPNQVLEIGDEVLEFECQANGHPRTLYHVSVEGNSILLPGT--RDGHEWTLTPE 355
 DB 317 sfgtkpadsqpagytatfectlvsgspayfvskeggdglfpyvsadgrtkvapt-- 374
 QY 356 GRSVLISARFADSDGKVTNCALNAVGVSSRTVSVSDTQEL----- 399
 DB 375 --gtltievrqvdgeyav-cagmasagsslsk--aalkatfetkgrvqkkskngkqkq 429
 QY 400 -----PPPIEQGPVNOQLPVKSVIVLPCRTLGTVPVQVSWYLD 438
 DB 430 knvqslkilylavtgaotkapkqptlieghbqntlmvgssallpcqasgkptgslwlrld 489
 QY 439 GPIVDQBERHNLSDAGALTSIDLQRHEDBSLYTCVSNRNGKSGWSGLDLPNPN 498
 DB 490 glpdlidts-srisqstgslhladlkk-pdgvtyctlaanedgestwsaalvdehts-n 546

QY 499 IKFFRAPELSTYPPPGPKQPMVEKGENSEVLSWTRSNKVGSSLVGVYIDMFKNMIDGW 558
 DB 547 aqfvrmpdpnsfespqtqilvntdtehlvnmastgeagptigtgytyspdlqgtw 606
 QY 559 VAVGTREVQNTITFTQGLLPGVNYFLIRAEHNSGLSLPMSPEFTVGTFRNSGL---- 614
 DB 607 fnldpvyasteyrlkpkshymfiraenektgtpessalvttskpaqvalsdkn 666
 QY 615 --LLSEARASILSGDGVLENSASVVDSTSMKLWQIIN-KTYVEGFTVIARQLNPPIVNN 671
 DB 667 kmdmaiaekrlitseqlikvktinstavrflwkrkrlceelldgyyikw----- 717
 QY 672 PAPVTSNTNPLGSTSTSSASASASALISTKPNIAAGKDEGKDEGSGGAPPTLNTKY 731
 DB 718 -gprtdndgqvntps----- 734
 QY 732 RMLTILNGGGASCTITGLVYTLFEFFVPTFK---SVBKGKPSNRRIARTEDVSEAP 788
 DB 735 -----tanyvsnlmpftnyeffivphsgvshshgapsndvltaaapslpp 784
 QY 789 YGMEALLANSANTVLAANKAPELKDHRGVLYNIVVVGIDTAHNFSILNVTIDAASTP 848
 DB 785 edvrimlnlttiriskwkapdgnglkgfgvlgv--qapnnr---nittneras 839
 QY 849 LVLANLTGCVMTVGVAGNAGVGPY--CVPALALDITKRL-----DPPIN---QR 897
 DB 840 vtlfvlvtmtykivraarsngyvgvstsevmldgtlekhlaqaagnefyglnk 899
 QY 898 DEVDWDLQVFWILLGAILVNLSPGAMVFK-----RKBNMKQSLANTHRCN 948
 DB 900 shvp-----vivvalilifvillaycywnsrnsdgkdrsfikladvsh-masn 950
 QY 949 HTSDWLKMS-----LSARNGNFWLSDST-----GHWWRPSPGSD 985
 DB 951 nlwvdaqngpnmyntagrtmmnrngalysltlnaagdfinnodsgytnhrps--- 1006
 QY 986 SLENQKCHADTAPVCGAGPSGAGGTSGGSGAGSAGGDDHGHGSGERNQRTVQ 1045
 DB 1007 ---sehhyhaqitgpgn--- 1022
 QY 1046 EYNSMIDTAEVSVFGKAPSEYGRGNASPAYATSSILSPHQ-----QOQQQPRITQ 1099
 DB 1023 -----amstf-----yngyhddpsyattvllnsqgpawndkmlrapmt 1066
 QY 1100 RPPVCTGLQRFHMP-----HYQQQHQQOQOQO-----THQHQALQHQQLPSPNI 1146
 DB 1067 nppv-----pepparyadhtagrsrsrsasdgrrtlogglhrtsgsrsdpshtdw 1120
 QY 1147 -YQMSITSEIYPTNTPGSRVSVSYQYYPKDKRQHRIHTENKL-----SN----- 1191
 DB 1121 syvghsd-----gtgskertgertpp-----nktlmdfipppppppppp 1151
 QY 1192 CHTYEAAGA--KQSSPSSQFAS-----VRRQLPNCSIGRESARFKVLTZ 1238
 DB 1165 ghyvdtatrrglrystpredtydsdsgafarvdwarpstsrnrlggprlkgk-rdd 1222
 QY 1239 QGKNQNLQDLDSGSMCYGLADSG-----CGSPSPMNLMSHREDHALTH 1285
 DB 1223 dgsrslmddggsdeagdegsedvprgvrkvaprgmlasltla-----hscy 1274
 QY 1286 T 1286
 DB 1275 t 1275

RESULT 8
 Y08403
 ID Y08403 standard; Protein; 1297 AA.
 XX
 AC Y08403;
 XX
 DT 24-JUL-1999 (first entry)
 XX

```
Qy      4 PRHIEHPDITVKNKDPFFPCQAENPPTIQWKDGRL-----KTDSGSHNIPAGGL 60
Db      30 prviehpdpvvsrgspatlnckgpc-takitvykqggvptlneqmsnshrivldtgsL 88

        61 PFLAKVHSR--RESDAQTYWCAEKFVGARSNKATLVANLRDFLEFANTVRVAQEVL 118
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      89 flfkwsxknkgksdageyyvcsahgehvkegnslklmLnredfrvrtpvtqayglam 148

Qy     119 ALMECGAPRGSPFPQISWKKNGQLTNLVGNKRIRIVDGGNLAIQARSGDDGRQCVCVN 178
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     149 avlescprrpfepvsvskddkelrldmqprythlsdgllldpvdrsdsgtyqvann 208

Qy     179 VVGTSRESATFLKHVRFFFLIRGPNQNTAVGSSVVSQCRIGDPLPOVLMRTASGGM 238
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     209 mgyversnaprlsvfekpfegpekdntdvgaavildcrtvdyppgitwkr---knem 266

Qy     239 PLRKFSMLHASGVHVLED-RSLKLDVVLTEDMGVEYCEADNAVGGITATGILTHAPP 297
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     267 pvt-----raylkdngrlriervpsdegeyyarnpstgesaahlrvqap 316

Qy     298 KEFVIRPNQVLVEIGDEVLFECQANGRPRLTVSVSEGNSSLPGY--RDGRMEVLTPE 355
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     317 sfgtkpdsgvpagtgatfecflvggspayfwkesqgdallfpseydsadgrtkvaspt- 374

Qy     356 GRSVLSTIAREFAREDSKVCTCNALNAVGVSRRTYSVDTPQEL----- 399
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     375 --gtltieevrgdegayv-cagmsagsksl--aalkafetkrvgkkssknkgkqk 429

Qy     400 -----PPPIBQQPVNQTLPKVSIYVLPCTRITVPVQWSVLD 438
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     430 knvssilkyvlisavtznatkpnctfehgantlmussalncocasektfnaiswrd 489
```

Qy	439	GIPIDQHEERHNRJSDGALITISDLQRHDEGLITCVASNKRNGSGSLYLRDTFNNF	498
Db	490	gipiditd-srisnsgtsgihadiak-pdtgytyctiaknedgsetwsalvdehts-n	546
Qy	499	IKFPAFELSITVPGPGKQWVKGSNVLSTWRSNKGSSGLVGYIEMKNETDGG	558
Db	547	aqfvrmpdpsfsspgplgvnvdtelvhwnagstagspigttyilyvyspdgigtw	606
Qy	559	VAGWTRWNTVPTOTGLLPGVNYPLIRADNSGSLSPMSPEITVGTRTNSGL----	614
Db	607	fnfdpdyasteytqglkphsgmyvliraenekigtstvasalvttspaagvdsdka	666
Qy	615	---DLSEARASLLSGDVNLSASVSDVSTSMKLTVQIIN-KYFVEGYVTARQLPNFVIN	671
Db	667	kmmdlaekrltsqglilevktinsstavrlfwkrrleelidgytykr-----	717
Qy	672	PAPVTSNTNPLGSTSTASASASALISTKPNIAAAGKROGETNQSGGGAPPLNTKY	731
Db	718	---gpprdnqgryvntsp-----	734
Qy	732	RMLTILNGSGASSCTITGLWVLYTYEFTYFFPK-----SVEKGPSNSRIARTLEDVSPAP	788
Db	735	---tenyvnslmptnyefvilyphsgshghagpansmdvtlaeapslpp	784
Qy	789	YGMELLANSVAFKLWAKELPKDRHGVLLNYHYIVRGIDTAHFSRILTNVTIDAASPT	848
Db	785	edvrlmlnlttlriswkapgdadingilkgfvlvg---gaponnr---nittreras	839
Qy	849	LVLANLTBGWNTVTVGAAGNAGVGPY-CVPATRLDPITKRL-----DPFIN---QR	897
Db	840	vtlflvltmytkyrlvaarsnggyvshgtsevmqldtlekhlhaaqenesfilyglnk	899
Qy	898	DHYNDVLQTFWILLGALLVLMGSGAMVFK-----RKHMOKQSALTNMRGN	948
Db	900	shwp-----vivlvalvilyfwilaycywnrsnsgdsrkfinkdngsh-masn	950
Qy	949	HTSDVLKMP-----LSARNGVYLDST-----GGMYWRPSGGD	985
Db	951	nldvqagpnaqpmnytagtzmnnrngalsltpnaqdffnodcstynhrp-----	1006
Qy	986	SLEMCKHDIAIDVAPVCAPGSPAGGETSSGSGAGSGASGSDGIDHGGHSERNQRTVG	1045
Db	1007	---sehyhyagltggpgn-----	1022
Qy	1046	EYSNIPTDTSVSSFGKAPSEYGRHGNASPAYATSSILSPHQ-----QQQQPFIYQ	1099
Db	1023	---msf-----yngyhdppsyattvlisnqcapawldnkirapamt	1066
Qy	1100	RPVPGYGLGRMHP-----HYOQOHOQOQOQ-----THQHOALQHQQLPSNI	1146
Db	1067	npvp-----pepparyadhtagrsrsrsdrgtllngllhrtsgsrgdsphdtv	1120
Qy	1147	---YQOMSTSTSEITTPGKRSYTSQYIYTKDKRHIHTENKL-----SW-----	1191
Db	1121	syvqlhssd-----gtgsskertzerttp-----nktlmdfipppnppppp	1164
Qy	1192	CHTYEAAGA---KQSPISSSQFAS-----VRQQLPPNCISGRSAREFVLMTD	1238
Db	1165	ghwydtatrrylngstpretdydsvsdgsafzvdvnartsnrnlgrrplkgk-rdd	1222
Qy	1239	QSKNQOQLNDLDSGSMYCNGLADSG-----CGSGSPAMLMSEDEHALYH	1285
Db	1223	dsgrslmmdidggsseadgensegdyvrggyrvkavpmglsastla-----hscy	1274
Qy	1286	T 1286	
Db	1275	T 1275	

RESULT 9
W83927
ID W83927 standard: Protein: 753 AA

```

RESULT 10
R13144
ID R13144 standard; Protein; 1728 AA.
XX
AC R13144;
XX

```

Qy	5	RIIEHPMTVPKNDPPTFCQAQGN-PTPIIWKQDKRELKDTGSHRIMLAGGFLIF	63
Db	242	fflspesdavatargnwlldcsaesgrvpyikwkdghlalmgderkqslsnglliq	301
Qy	64	KVHSR-RESAGTCTKCEA-NEFGVARNRATLQVA-VLDRFEPLEPANTVRAGVEAL	120
Db	302	nlhsrhhkpedqlycsealsdgsyisrtakvavaplr-flsqsevatfngdvtl	359
Qy	121	MEGAGPRGSPQFQISWRKNGNLN-LWGNKRIRIVDGNLAIQEARQSDGRTQCVVNY	179
Db	360	lkecv-igepmthlwnkqgdqitppdgsrvllpsgalisrlpdpqiyircsarp	418
Qy	180	VGTREASTAFILKV-----HYRFFLLINGPQNGTAVGSSVYFCRIGDPLDPLWRRITA	233
Db	419	asrtngeawmlsdpghrlhrlpssvalseagkdavlecsygyppsfatglr	478
Qy	234	SCGNPLAKFESWILASGRGVWLEDRSLKLDSTVQCTGCTEADNVAWGTATIGLE	293
Db	479	evqlrskrys-----llqgnallsivtdddsgmytcvtykneiasaelvt	528
Qy	294	HAPPKVFQKNQVLEIGVEIDFQALGCPHPTIYLVSGNENSLLLPYRGDMREWTLT	353

Db 529 lvpvflmipenlyayesmdiefectvsgkpvptvnmkngd -vvip-----sdyfqiv- 581
QY 354 PEGRSVLIARFAREDSGKVTWCNALNVGVSRSVTVQDFELPPPIEQGPVNTL 413
Db 582 -ggsnrlrlgylwksdge-fygcvaeneagngt-----saqlrvpkalpssvlpa 632
QY 414 PVKSVLTPCRLTGTPVPOVSXNLGIPDV-----QEHHER-----NLSAGA 457
Db 633 prdvrvpl-----vsrfvrlw-----rppeakngiqtffvfrsregdnreralntqgs 685
QY 458 -LITSDQREHDEGLIT--CVASNRNKSWSGYSRLDPTPNMIAKFAPELSTYGP 513
Db 686 lqltvgnl--kpeamrtfrrvaynewgpg-----sspklvatqpelgvpgp 73
QY 514 PGKPMOYKGENSVLTVTRSMKVGSSSLVCFVME-----GKMETDGWANGTVQN 567
Db 732 venlqavstpslilicwepgyang-pvgy--rlfctevstqeg-----niewd 781
QY 568 TITPTQILGLRNVFFLIRASNKLSLSPMSEPIVGTFRVNSGLDSEARSLSD 627
Db 782 lsykleglkfkteysrlflaynry--pgvsddltvt-----ldsvpaggpn 830
QY 628 VVELSNASVDSITSMKLW-----QIINGXVZGFVV----- 659
Db 831 slc-----vnsrsikvslpppsqtng-fitykirkhrttgrmetlepnlywf 884
QY 660 -----YARQLPFIWNVNP-----VSNVT--NPL-----LGSSTSAASASALIS 701
Db 885 tglekgsyfygsavnttpgsnvtaetpendidesvvpdpsslhvrpqtncim 944
QY 702 T-----KPIAAAGKRG-----ET-----NQG 720
Db 945 swtpeplnrvirvgylgygygyvpyaetrvdsqkryysierlesshyvislkafng 1004
QY 721 GGAP-----|||-----TPL----- 727
Db 1005 egyplesatrsitdtpdvdypllddftvpdtpdtpmppvgvavalthdvrrs 1064
QY 728 -----NTRYMLII--LNGGAS-----SCTIGLWQYTLIEFFVIF 763
Db 1065 wadnsvpkntksevrltyvtrvtfassakysedtslytqtkpntvysfvmv 1124
QY 764 YKSVBKPFSNRIARTLEDVPSAPYGMALLNS--AVFLKNKAPELDRGVLNTH 821
Db 1125 knrrstwsmsahatyeaaptspakdfvtviregkpraviswqp--leangkitayl 1182
QY 822 VIVRGDIAHNFSLRLN-----TIDASPTVLIANLIEGVMTVVGACNANGV 873
Db 1183 1-----fytdknlpiddwmetisgdrlghimldntayfrlgarskvy 1232
QY 874 PYCP--ATLRLO-----PITRL-----DPFIORNDHVDVLR 905
Db 1233 plsdpllfrtkvehpdkmandgahgdgyvpydntldrstlelpgghpmhpgsvt 1292
QY 906 OP-----WPITLGAAILVLMSPGAMVFKKHMOKQSAALTMRGNH--TSDVAMP 957
Db 1293 pqknsnllfytlvtyvtyvvivavicttrssaggrkkrthsagrksgskldrp 1352
QY 958 SLSARGNGVYLDSTGGM--VWRPS--PGGDSLMQKHDIATVAPCGAPGSPAGGT 1012
Db 1353 di-----vhhemenkklspsgdpagdrspigs--cqdltpvshsgsetqlsg 1402
QY 1013 SSGSGG-----GAGSAGSGDDIGHGSGSERNQORTVGBYSNPTYD-----AYSFSGKA 1663
Db 1403 sthsgdteegaeasms-----tlerslaarprkl-----nmpdaqsnmpavswlp 1455
QY 1064 PSEYGRHNASPAYATSYSLSPHQOQQOQPYQYQRPV-----GYGLORPMHPHY 1115
Db 1456 tlesaypggilsp-----tcgyh-----poflrvpfpvptlsdrqfaggr----- 1498
QY 1116 QQQOQQOQQAQTHQOHAQHQOHLPPSNIQMSITSEIYPTNTPGSRV 1167
Db 1499 -----gsyveapttaopnpls-----qephssseasortlptacy 1537

RESULT 11
 R68553
 ID R68553 standard; Protein; 1447 AA.
 XX
 AC R68553;
 XX
 DT 05-JUL-1995 (first entry)
 XX
 DE Deleted in colorectal carcinoma (DCC).
 XX
 KW Tumour suppressor; deleted in colorectal carcinoma; antibody;
 KW cancer diagnosis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..1063
 FT /note= "DCC epitope"
 FT Misc-difference 3369..4341
 FT /note= "DCC epitope"
 FT Misc-difference 26..1126
 FT /note= "DCC epitope on extracellular domain"
 FT Misc-difference 1123..1447
 FT /note= "DCC epitope on intracellular domain"
 XX
 PN W09428161-A.
 PD 08-DEC-1994.
 XX
 PF 18-MAY-1994; 94WO-US05277.
 XX
 PR 26-MAY-1993; 93US-0068950.
 XX
 PA (UJJO) UNIV JOHNS HOPKINS.
 XX
 PI Bruskin A, Jarosz DE, Johnson K, Kinzler KW, Vogelstein B;
 PI Labrecky JR;
 XX
 DR WPI; 1995-022830/03.
 DR P-PSDB; Q80196.
 XX
 PT Antibodies specific for tumour suppressor gene product, DCC -
 PT useful for detecting expression of DCC gene, for cancer diagnosis.
 XX
 PS Claim 4; Page 24-28; 39pp; English.
 XX
 CC The protein represents the DCC tumor suppressor, and epitopes are
 CC identified which are used in the generation of polyclonal or
 CC or monoclonal antibodies against DCC. The antibodies can detect
 CC DCC protein in biological samples (including tumour tissue,
 CC peripheral blood mononuclear cells or a tumour biopsy lysate)
 CC despite low levels of DCC expression, and are therefore useful in
 CC cancer, especially colorectal carcinoma, diagnosis.
 XX
 SQ Sequence 1447 AA;

 Query Match 7.6%; Score 551.5; DB 16; Length 1447;
 Best Local Similarity 21.8%; Pred. No. 1.3e-25;
 Matches 312; Conservative 177; Mismatches 542; Indels 397; Gaps 64;

 QY 5 RIIEHPMDTTPKNDPPFFNCAQBN-PTPIQWFKDRELKTDGSHRIMLAGLFFL 63
 DB 41 rfllsepsdavntrngvllcdsaesdrgygvikwkdghlaldmderkqqlngslliq 100

 QY 64 KVIHSR-RESAGTWCRAK-NEFCVARSRNATLQVA-VLRDEFLRSPANTRWAGYVAL 120
 DB 101 nlhshrhkpegelqycoasldgdsqslsrktakvavagplr-flsqtesvafmgdtrvl 158

 QY 121 MECCAPGSGSPPOISWRKNGQTLN-LVGNKRIRIVGSGALQIARQSDGRQYKVN 179
 DB 1152 dl-----wihheememknekpsgtodagrdspiqs--cqdltpshsgetqlgsk 1201

Db 159 lkcew-lgepmthlwkgqkdtltpgdsrvvpslqslrsgpdgilyrcsarp 217
 QY 180 VGTRESATAFIKV-----HVPFLIRGPNQAVVSGSVVFCRIGDGLDVLAKRTA 233
 Db 218 asrtqneaevrllsdgplhrqlyflrpsnvaleqkdavleocvesgypppsfwlrge 277
 QY 234 SGGNPLRKFSLNLSASGRVHVLDESLKLDVLEDWGEYTCRADNAGVGTATGILTV 293
 Db 278 evrlqrlkyskys-----llgssnllsvntddsgmytcvtykneisaaseltv 327
 QY 294 HAPPKTVIRPNQVLVEIGDEVLFECQNGHPRLPLYNVEGNSLLGLCYRGRMEVILT 353
 Db 328 lvppwflnhsnlyayesmdiefectvsgkpyptvnmknqd--vvip--sdyfqlv-- 380
 QY 354 PEGRSVLSIARFARSDSGKVVTCNLNANGSVSSRTVSVSDTQELPPLPIEQGPNQTL 413
 Db 381 --ggsnlrlilgyvksdeg-fyqcvaeneagngt-----saqlivpkpaipssvlpea 431
 QY 414 PVLSIPLWCHLTGTPVPSVWYLDGIPDV-----QEHRR-----NLSAGA 457
 Db 432 prdvwpvl-----vssrfvlsw--rppeaknglqtfvtfvregdnrealttqgs 484
 QY 458 --LTIISLQHEDEGLYLT--CVASNRNKSNSGSLRLDTFNFNIRKFAPELSTYGP 513
 Db 485 lqltvgnl--Apeaytfrrvaywepgpe-----ssqplvatqelqv-pgp 530
 QY 514 PGKPNQVKGNSVTLSTRSKMKVGGSLVGIYEMF-----GKNETDGNWAVTRQVN 567
 Db 531 venlqavstpsliltwepayang-pvqgy--rlfctevstgkeg-----nlerdy 580
 QY 568 TTTPTGGLPGLVNNFFLIRAEHSHLSLPSMSEPITVGRFYNSGLDSEARASLLSG 627
 Db 581 lsyklleglkkfteyslrlfayrny--pygstddltvt-----lsdpsappgnv 629
 QY 628 NVELSNASVSDTSSMKLT-----QIINKYVGEFVY----- 659
 Db 630 sle-----vnsrslkvsllppsgtqg-fitygkirkhrttrrgemetlennlylf 683
 QY 660 -----YAKLPNIPVNPAP--VTSMT--NPL-----LGSSTASASASALIS 701
 Db 684 tgleksgysqsyfgeantvngtppsnvtaetpendidesqydgslhvrptgclm 743
 QY 702 T-----KPIIAAKKRG-----ET-----NQSG 720
 Db 744 swtppnlnvrrvgyllygyvgsyapavrvdsqkrqyslerlessshyislakfnag 803
 QY 721 GGAP-----TFL----- 727
 Db 804 egyplyseattrstldtcdpvdypplldfptsavplstpmppvgvqavalthdavr 863
 QY 728 -----NTRYKMLTI--LNGGAS-----SCTIIGLQVLTLYEFIVFF 763
 Db 864 wadnsyvkqkqteevlrvtrvtsfasakysedttlsyatlakptnqfysvmt 923
 QY 764 YKSVDPKPSNRRIARLDDVPSRAPTYGMEALLNS--AVFLKAPELKDRHGVLLNTH 821
 Db 924 knrrstwmstahattyaaptspakdfvtviregkpravivswpp--leangkitayl 981
 QY 822 VIVRGIDTAHNSRIILTV-----TIDAASPTVLNLTIGWMTVTVGAAGNAGW 873
 Db 982 l-----fytdknlpiddwmetisgdrltbgldmldtmyfirqnarskyg 1031
 QY 874 PTCVP--ATILRL-----PITRKL-----DPFINQRHGVNDVLT 905
 Db 1032 plsdplrlfrtkvhepdkmandgrghdggyvrvdnlndrlstneppignmpghpsvt 1091
 QY 906 QP-----WFILIGLALVNLSPGAVYVKRKHMMQSQALMTWNGNN--TSDVLKWP 957
 Db 1092 pknsnllvlrvvtrvtrvvlvavictrrsaqgkkratkhesgkrkgsgdlrpp 1151
 QY 958 SLASRNGNGYKLDSSTGM--VWRPS--PGDSLQMKNDIADYAVPCGAPSGAGSGT 1012
 Db 1152 dl-----wihheememknekpsgtodagrdspiqs--cqdltpshsgetqlgsk 1201

A... A&A HETEROGENEOUS POLYMERIZATION OF BUTADIENE-1,3

Db 962fytlidknipiddwlnetisgdrthqimldmtyvfrigaraskvg 1031
 Qy 874 PYCPV---ATLRLD-----PITKRL-----DPFINQQRHVDVLT 905
 Db 1032 plsdplfrlfrlkvehpdkmandqgrhgdgyvvpdtnldrstleppigqmhphgsvt 1091
 Qy 906 QP-----WFIIIGAILAVLMLPGSNAVFKYKRHMNMQSALNMRGNE--TSDVLMKP 957
 Db 1092 pqknsallviivtvgvvtlvvviavictrrssaqgrkkkrathssagkrksgkdrlpp 1151
 Qy 958 SLARMGNWGLVDSSTGM---WVRPS---PGGSLMQRDHIADYAPVCGAPSGAGGT 1012
 Db 1152 dl-----wihheememknlekgsgtdpgrdpsigs---cpdltpvshsqsqtqlsg 1201
 Qy 1013 SSGSGS-----GAGSGAGGGDDLHGHHGSENRQRY-VGEYSNIPDTEVYSSFGKAPSEY 1067
 Db 1202 stshsgqddteagssmstlerslaarapraklmipmdaggnpp---avvsalpvptles 1258
 Qy 1068 GRHGNASFPAYTSSILSPHQOQQOQPPRYOORPV-----GYGLQRMHPHYQOQ 1119
 Db 1259 aayglplsp---tcgyph-----pgfllrpfvftlsvdrgfag-- 1297
 Qy 1120 HQOQAQOQTHQOHALQOQLOPPSNYQOMSTSEYTPNTGSPRSV 1167
 Db 1298 -----sqvsegpttqgpnlpss---qphssseapartipiacv 1336

RESULT 13

W57900
 ID W57900 standard; Protein; 1192 AA.
 XX
 AC W57900;
 DT 25-SEP-1998 (first entry)
 DE Protein of clone C0722_1.
 XX
 KW Human; nutritional supplement; cell proliferation/differentiation;
 KW cytokine; immunostimulant; immunosuppressant; haematopoiesis regulator;
 KW receptor/ligand activity; cadherin/tumour invasion suppressor;
 KW anti-inflammatory; tumour inhibitor; clone C0722_1.
 XX
 OS Homo sapiens.
 PW
 PX W09824905-A2.
 XX
 PD 11-JUN-1998.
 XX
 PR 05-DEC-1997; 97W0-US22211.
 XX
 PR 03-DEC-1997; 97US-0984516.
 PR 06-DEC-1996; 96US-0762216.
 PA (GEMT) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavelle ER, McCoy JW, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI; 1998-333324/29.
 DR N-PSDB; V40887.
 XX
 PT New isolated polynucleotides encoding secreted polypeptides -
 PT isolated from a human foetal kidney cDNA library, a human adult
 PT blood cDNA library or a human adult brain cDNA library
 XX
 PS Claim 36; Page 81-85; 10pp; English.
 XX
 CC This sequence represents the protein of clone C0722_1, of the
 CC invention. DNA encoding this sequence was isolated from a human adult
 CC brain cDNA library. The DNAs and proteins can be used as nutritional
 CC sources or supplements, or may exhibit cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or suppressing

CC activity, haematopoiesis regulating activity, receptor/ligand activity,
 CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
 CC tumour inhibition activity or other activities.

XX
 SQ Sequence 1192 AA;

Query Match 7.5%; Score 548.5; DB 19; Length 1192;
 Best Local Similarity 21.2%; Pref. No. 1.5e-25;
 Matches 273; Conservative 157; Mismatches 510; Indels 345; Gaps 48;
 Qy 3 NPRIIE-----HPMDTTPVKNPPTFMQAGNPTPTQWKGDLGRLATDGS 51
 Db 35 dkdilledvqptitqgskdyldiprenivigeceakgpppsfwtngthfdidkpl 94
 Qy 52 RIMLPAGGLFLVLI-HSRRES DAGTYWCBAKNGFVARSNATLQV--AVLDEFLPE 108
 Db 95 vtnkgptgtllinimsegkaetyegvyqctarnergaavsnlvrvpsrplwtkelep 154
 Qy 109 ANTRVAGQVGLMCEGAPKSGPEPQISWIKNG----- 140
 Db 155 it--lqsgslvipcrpplgplpplifmdnsfqlpqservsqglngldlyfswlpd 212
 Qy 141 -----QTL-----NLYGNKRIR----- 152
 Db 213 redyicyarfhtqitqgkplsvkviskssrrptftlpegnasneklrgvlsle 272
 Qy 153 -IVDG-----GNLAIQEARQSDGRGVQKVVKNVGTRES 185
 Db 273 claepltplywakedgmlkrtvnykfektlqihlseadsgnygcknalgaih 332
 Qy 186 ATAFILVHVHVFILRGPOQTAVNGSSVFEQCRIGDPLVDLWRRFASGGMNFKFSW 243
 Db 333 -tlsvrkaepwyitapgnvlspgedgtlcrangnkp-----rlsw 375
 Qy 246 LHASGVRVHVEDSLKLDVTL-----EDMGVTCADNAVGGITATGTLVHAPK 298
 Db 376 ltagvpleapdpgkldgtlilsanvqersaavqcnasneyllanefvnlleap 435
 Qy 299 FVIRKPNQLVIE--GDEVFECAQHGRPRFLTVSEVGNSSLLGLP---YDRGMEVLT 352
 Db 436 riltpantlyqianrpaldcaffsgpltlewfkagksalhediyhengtlelep 495
 Qy 353 TPBGSVLSIARFARSDGKVTCNALNAVGSVSRVTVS-DTQFELPPP---IIEGQ 408
 Db 496 -----aqkdstgtctvarknlgmaknevhlkaidatvkwqeyavrg- 541
 Qy 409 VNQTLVKVSIVLPCR---TLGTPVPQVSYLDG--IPIDVQHERNNISDAGALIS 461
 Db 542 -----smvsfeckvkhdtlsl---tlvldkdrelpsd-----erftvdkhvlsa 585
 Qy 462 DLQHRDEBLGTLVCSANNKGSWSGSLRLDPT-NPNKIFFRAPLSTTPGPKQPMV 520
 Db 586 dve-dddsgtctvamtldsvsavsavtaptcp-----ap-vydvgnppfdelt 637
 Qy 521 EKGENVSLVLSMTRSNKVGSSLVGYVEMKGTDMGVAVGTQVNT-TFTQTLGGV 579
 Db 638 dqlkdsqvlsvtggdd-nspkltkfiedamhpgplvhhgtetsvtgtatqlklsprv 696
 Qy 580 NYFFLIARNHSGLSLPSMSEPTVGTVYFNSGLDSEARASLSGSDVSNLSNVSDS 639
 Db 697 nysfvmavnsigklspease-----gylkasepdk-----nptaveg 736
 Qy 640 -----TSMKLTWQINGIKYEC---FYVYARLPNIPVNPVPTNTNPLGSGTSTA 690
 Db 737 lseepdlvltwplngfesngplgykvsvrqkdgdewtsvvanvsklysvgtptfv 796
 Qy 691 SASASASAL--ISTKPNIAAACKROGE-----TNQSGGGAAPLNT-- 729
 Db 797 pylirkvqindcgaapepavnnhgedlpmvavpgrvrvnnstlaewhwdpklksr 856
 Qy 730 -----KYRMLTILANGSGASCTTIGLWQTYLFEFFVFPYK 765

Query Match 7.5%; Score 543.5; DB 19; Length 1028;
Best Local Similarity 23.0%; Pred. No. 2.4e-25;
Matches 263; Conservative 124; Mismatches 425; Indels 333; Gaps 44;

Qy 4 PRIIEHFNHPTVP-----KNDPFFVHCQABGNPTPTTQWCKDGRLEKTDGSGHMLPAGOL 60
Db 26 piftgqehpvdhifpdlidskseviloacaangspphyrvkqngtdidf--tmsyhyldrgsl 84
Qy 61 FFLKVIHS--RESAGDTWCKEAKNEFVGASRNATQVAVLRDPEFLPAMTVAGQVGA 119
Db 85 -----ainsphtddigmyqlcatnllgtlrsakligfayedfekttrstvsveggvg 140
Qy 120 LMECKAPGSGPEPQISWRNKQJTLNL-VGNKAIRVVDGMLAIEQARQSDGGRQCVKWN 178
Db 141 vllqppppghfdgysawtndnplvlgvdnrrfvsqegnliakpvedsnyntcfctn 200
Qy 179 VVGTS-----ESATFLKVVHVPFRLIGQPNQTVAGVSSVVFQCRIGGDPPLDVL 258
Db 201 keasrgsvggpttvlvrgtdgmygeyepkleivrfpetlpaqadskvklecfalgpnvpid 260
Qy 229 WRRTASGGMPLKFLKSWLSASRGVHVEDRS--LLKDDVLEDMGETTCEADNAGVGITA 287
Db 261 wr--ldgsp-----gkykysqallelpgfegedegfyeciasnryrnlga 308
Qy 288 TGLITVHAPPKVFRPNKQJVEIGDEVLFECQANGHPRPTFLVTSVSG-----N 335
Db 309 kqglfypawewegklqnlthlsyldnlweckasxkpnpywtvkwgerlnpeerligien 368
Qy 336 SLLLLPG-----YDQGMVEY----- 350
Db 369 gltliitmvnsdgyvgvcaenkygilyanaelrvilasapdfksvkvkksfvvgvgdv 428
Qy 351 -----TLTFEGS-----SVLSIARFAREDSKVVTCKNALMAGVS 384
Db 429 igckpnafpraalswrgkrtetlrgskriflledgsiklynltrsdags-tyclatnftgt 487
Qy 385 -----VSSRTVSVSDTQFLPPEIIEQGVNQTLPVKVISVLPCLRTLTPVPQV--S 434
Db 488 akntgsilvkertvtp-----psmdvtvgesilpqchshdpdeservf 534
Qy 435 WYLDGIPID-----VQEHERRNLSDAGALTISDLQRHEDGLTVCVSNRNGKSSNGYL 490
Db 535 Wfngdvlidkgyahferlgesesvgdmlnrlqhlh--skgylctvqt----- 581
Qy 491 LDTPTNPMKIFRPAPELSTYPPGPKQWKEGSENVY-----LSWTRSNKVGSSLVGYVI 547
Db 582 -----tleslavadiivrgpppggedqvedisstqslw-ragpdnspqiflti 633
Qy 548 EMFGKETD--GWAAGV-----RVQNTPTTQGLLPGWYTFFLIARNSHGLSLPS 597
Db 634 q-----trtpfsgvgevtpellingjyativ--vglsmpweyefrvagnsiglges 687
Qy 598 PMSE-----PIITGTRFNSLDLSEARASLSGGVVELSNASVQVTSMKLTWQII 649
Db 688 epeelrtkxasvpgvapvnlhggysgrse-----lyvtwesi 724
Qy 650 -----NGKTVGFG-VYIARQLPNVFNPNVPVTSNPNLLGSTSTAS--ASASALIS 701
Db 725 peelng--egfyf-----lnfrp-----vgstwskevsvessvrfy 763
Qy 702 TKPNIAAAGKRDGTNVSOGGGAFTPLNTKMYLTILNGQASASTTGTGWLQTYLYEFFV 761
Db 764 rnesi-----lpslpfeykvgyneq----- 785
Qy 762 PFTYSVEGKPNRSNIARTLEDYVSEAPYGEALLNSAVFLKMAPELKDHRHVLNHY 821
Db 786 -----egslvtvtvysgedepqlargtsglsasenevsnaiamntvrglye 829
Qy 822 VIVRGIDTAHNF-----SKRLNVTIDAASPTVLANLEGVMTVGVVAGNAGVGYPCVP 878
Db 840 vlywtddsmesiklgirsgvnt-----kntgtikantlyfasvryntagtcgsspp 893
Qy 879 ATLRLDPTPKLDPFPIQRDHFVNDVLTQ-----WFI-----ILLGAILAVMLSLGAMVF 929
Db 894 vnv-----ltkssp-----sgppanawlknstkslnclnwhkvtkenesevlg 937
Qy 930 VKKKEDMOKSALNTWRGNHTSDVLEKPLSARNGNWTNDSSTGMYWRPSPGSDGLE 989

Db 938 ykilyrgqrgskthilemtntsaellvpf-----eedyliert-----vsdggdgses 986

Qy 990 QKDHI 994

Db 987 eelri 991

RESULT 15

W42087

ID W42087 standard; Protein; 1571 AA.

XX

AC W42087;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM2.

XX

KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;

KW signal transduction; trisomy 21; mental retardation;

KW holoprosencephaly; corpus callosum agenesis;

KW schizencephaly; diagnosis; assay; human.

GS Homo sapiens.

XX

PN W09817755-A1.

XX

PD 30-APR-1998.

XX

PF 23-OCT-1997; 97WO-0519547.

XX

PR 25-OCT-1996; 96DS-0029322.

XX

PA (NCDA-) CEDARS SINAI MEDICAL CENT.

XX

PI Korenberg JR;

XX

DR WPI: 1998-271791/24.

XX

DR N-PSDB; W31908.

XX

PT New isolated Down's Syndrome-cell adhesion molecule - used to

PT develop products for detection, diagnosis and therapy of

PT developmental and neurological abnormalities

XX

PS Claim 2; Page 90-95; 109pp; English.

XX

CC This polypeptide comprises Down syndrome-cell adhesion molecule
CC DS-CAM2, an extracellular soluble protein belonging to a novel
CC subclass of the Ig superfamily with highest homology to neural cell
CC adhesion molecules. Its amino acid sequence was deduced from cDNA
CC clones (see V31362) isolated from a trisomy 21 foetal brain library.
CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and
CC lacks the entire transmembrane domain of DS-CAM1. The invention
CC provides human and murine DS-CAM nucleic acid sequences (see also
CC V31361, V31365-87), expression vectors and host cells, transgenic
CC animals, antibodies, antisense oligonucleotides, and primers
CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
CC with developmental and neurological processes. They can be used in
CC e.g. neural prosthetic devices used in entubulation methods of
CC repairing (regenerating) damaged or severed peripheral nerves,
CC also in bioassays to identify agonists and antagonists. The products
CC can also be used in detection, diagnosis and therapy of developmental
CC and neurological abnormalities such as Down syndrome, mental
CC retardation, holoprosencephaly, agenesis of the corpus callosum,
CC or schizencephaly.

XX

SQ Sequence 1571 AA;

Query Match 7.4%; Score 538; DB 19; Length 1571;
Best Local Similarity 21.7%; Pred. No. 9.5e-25;
Matches 286; Conservative 166; Mismatches 458; Indels 410; Gaps 55;

Qy 25 QJAEENPTPTQWFKD-----GRELKTDTGSHRMLPAGGLFPLKVLHSRRSDAGTY 77

Db 246 kcalghepdyrwldkmplesgrgqktvg-----llienirpsdgsy 291

Qy 78 WCEAKNEFGVARSKNATLQVAVLRDEFLEPANTFVQGEVALMECGAPGSGPEQISWR 137

Db 292 vcevsarygtakvigrlykplk-atlsprkvsqsgvslscsv-tgtdedqlsavy 348

Qy 138 KNGQTLNLVGNKRIIVDGGNLAIQEARSDGRTQCVK-----177

Db 349 ngeillppgkvrwritglnhenlmdhwmksdggayqcfvrkdklsaqdyvqvrltdgtpk 408

Qy 178 -----NVVGF-----182

Db 409 tisafsekvsvpaepvslmncvkgtpplptitwtldddplksgshrismitsegnvisy 468

Qy 183 -----RESATAPL---KVHVR-PFLIRPQNGTAVVGSVVFQCRIG 220

Db 469 lnissqrdgdyvrrctannsgvlyqarinvrsgasirpmknkltalagrdtyihcrvi 528

Qy 221 GDLPLQVLAHRTAS-----234

Db 529 gypysyikwyksnallpfhnrqafenngtklisdvqkevdegeytcnvlqplstsgs 588

Qy 235 -----GCNPLRKFSLHSA---SGRVHLEDR 259

Db 589 vhtvkvppfipqfefrpfisgrvrfpcvsvgdlpi-titwkdgrplpgslgvtldn 647

Qy 260 -----SLKLDVLTLEMGEYTCEDANVGGTATGILTVHAPKPIVIRPQNLVEIGDEV 314

Db 648 ldfstslrlslslmhagnytlclarneaaahvsgqlrvrvppkfvrvpddgdygkav 707

Qy 315 LFDQALGHPPTLYLWSEGNSSL---LLPGKRDGHEWILTPEGRSVLSIARPAEDSGK 372

Db 708 lincseagypvptlrvfksagvpgfqlnigrvll---sngslilkhvreesdy 763

Qy 373 WTCNLANVGS-VSSRTVSVSVTQFELPPPIRQSPVMTLPVK-SIVVLPORTLTGPV 430

Db 764 yl-kyvnsdvadsksnlytkvl---pamtisy-pntlatgggkksctahgekp 816

Qy 431 PQVSWTLDGPIVDVQERR-----NLSDAAGLTISDLO---RHEDGTYTCV 475

Db 817 lfvrv-----ekedrlinpmarlyvstkevseevistqlilptvredsgffsch 866

Qy 476 ASNENKSSNGSLKRLDTPPNPKIKFFRAPELSTYPPGPKQWENG-ENSVLSTWR 533

Db 867 alnsydeg-rjiql-----tvgeppdppeleikdvartilrlwm 907

Qy 534 SNKYGSSSLNGVTLVEMGKNETDGWA-----VGTVRVMTFTTQTLGGLVNWFFLIRA 587

Db 908 gfd-gaspitgdylic-kksdwdasqtrkdvpsqlasatld---lhpstyslrmya 962

Qy 588 ENHSLGLSPSPSEPTVGTGTVNSFGDLSERASLLSGDVLSNASVDSMTKLWQ 647

Db 963 knrigsep---sneltl-----tadeagdpqpgv-hlepisssgrtwtk 1006

Qy 648 -----IINKTVEGFTYARQLP---NPVKNPAPVTS-----NTNP----LGS 685

Db 1007 apkhqlng-lirgylgyreystgngfgnlisvdtsgdsevytdlnakftaylvvq 1065

Qy 686 TSTSASASASALISTKPIAAAGKRGDGETNSGGGAPLMTKYRMLT-IUNG----739

Db 1066 acnragrtgssgelitt-tledvpsypnenqaiatspesiswslskealqllg 1123

Qy 740 -----GASSTCTT-----GLWQYTLVEFFVYFVSKBGKPSNRIR 778

Db 1124 frviyvanlmdgelgelknittqpsltdgleyktnysigvlatragdvrvsegiftr 1183

Qy 779 TLEDVSEAPTGMKALLNSSAVFLKNAPELKHGHVLLNHYVIRGIDTAHNSRLIT 838

Db 1184 tledvpg-ppgkkaaaasamvfw-llpkl-igilrkytvtf-----csbpyptvis 1235

Qy 839 NVITDASPTPLVLANLTGVMYTVGAAGNAGVGCVPATLRIDPDK-----RLDPFIN 895

| : : ||: | : | | : || | : : | : | : |
Db 1235 efeaspsdfsyrignlsrnrqysvwwvavtsagrgn--sseittveplakapariltf-- 1291
Oy 896 QRDHYNDVLTQPFIFILLGAILAVLMLSGAMVFVKRKHQKQSAINTMGHNSDVLK 955
: : | || : : | : | | | : :
Db 1292 -----sgtvttpw-----mkdivlpcakvgdgpavkwnkds-----ngtpslvt 1331
Oy 956 MPSELARNGNGTWL-----DSSTOGMYWRPSPGCDs--LEMQKDHIADTAPVCGAPGS 1006
: : || : : || : : | | : | : |
Db 1332 idgrsrlfsgsfliirtvkaedsgyyyclannnwgseilnlqgvppd-----qprl 1385
Oy 1007 PAGGGTSSGGSGGAGSGASGGDDIHGGHSSERNQQRVGEYSNIPTDAEVSSSGKAPSE 1066
||| : | : | | | : : || : : | : |||
Db 1386 tvskttssitlswlpgdngssirg-----yllqysednse--qwsfpispse 1433

Search completed: January 22, 2001, 12:17:58

Job time: 1635 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:24:13 ; Search time 325.28 Seconds
(without alignments)
288.277 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIHEPMDTTPVKNDP.....RSLLSNSGSGTSSQPAHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : FIR_56*
1: pir1*
2: pir2*
3: pir3*
4: pir4*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1489.5	20.5	1612	2 T30805	duttl protein - mo
2	1455.5	20.0	1651	2 T14160	transmembrane rece
3	1361.5	18.7	1273	2 T42405	sax-3 protein - Ca
4	1256	17.3	1344	2 T14316	rig-1 protein - mo
5	717	9.9	423	2 T29549	hypothetical prote
6	658	9.0	874	2 T29548	hypothetical prote
7	602	8.3	1028	2 T58164	BIG-1 protein - ra
8	595.5	8.2	1427	2 T51669	tumor suppressor -
9	593	8.2	1232	2 T43027	neural cell adhesi
10	581	8.0	1028	2 A53449	neurofascin - chic
11	579.5	8.0	1277	2 T30532	neural cell adhesi
12	568	7.8	1036	2 S22383	axonin 1 precursor
13	566	7.8	1259	2 A43425	Bravo/Nr-CAM cell
14	565.5	7.8	1272	2 S26180	plasmacytoma-assoc
15	560	7.7	1443	2 T50600	neogenin - chicken
16	558.5	7.7	1268	2 A39640	neural cell adhesi
17	551.5	7.6	1447	2 A54100	tumor suppressor p
18	544.5	7.5	1040	2 A49356	transient axonal g
19	543.5	7.5	1018	2 J04211	neural adhesion pr
20	543	7.5	1197	2 T30581	neural cell adhesi
21	540.5	7.4	1040	2 A34695	axonal glycoprotei
22	538	7.4	1896	2 T08851	Down syndrome cel
23	536.5	7.4	1018	2 A54744	contactin 1 precu
24	531.5	7.3	1020	2 S05944	neural cell surf
25	529	7.3	1239	1 A32579	neuroglian - fruit
26	525.5	7.2	6642	2 T29757	protein UNC-69 - C
27	524.5	7.2	1021	2 A57112	contactin precursor
28	521	7.2	2029	1 TDFFLK	protein-tyrosine-p
29	517.5	7.1	1010	2 J00094	F11 protein precu

30	517	7.1	1907	2 S50893	protein-tyrosine-p
31	516.5	7.1	1091	2 S01998	contactin precursor
32	511.5	7.0	1260	1 S05479	neural cell adhesi
33	509	7.0	1257	1 A41060	neural cell adhesi
34	497.5	6.8	1259	2 S36126	neural cell adhesi
35	495.5	6.8	1897	1 T08HLK	leukocyte antigen-
36	487	6.7	1894	2 C54689	protein-tyrosine-p
37	486	6.7	1898	2 A46216	leukocyte antigen-
38	486	6.7	5175	2 T20992	hypothetical prote
39	485.5	6.7	1912	2 A56178	protein-tyrosine-p
40	484	6.7	5198	2 T43290	hemocytin precurs
41	478	6.6	2783	2 T34416	hypothetical prote
42	476.5	6.6	1265	1 A37967	neural cell adhesi
43	475.5	6.5	2222	2 T13924	sdh protein - frui
44	475	6.5	1256	2 T03096	CDO protein - rat
45	461.5	6.3	1863	2 S46217	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T30805
duttl protein - mouse
N:Alternate names: transmembrane receptor protein Robol homolog
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30805
R:Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.
submitted to the EMBL Data Library, July 1998
A:Description: The mouse homologue of human DUTTL1/9-robol gene: protein sequence :
A:Reference number: E20879
A:Accession: T30805
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1612 <W6>
A:Cross-references: EMBL: T17793; NID: e1329712; PID: e1329713; PIDN: CAA76850.1
A:Experimental source: brain
C:Genetics:
A:Gene: duttl
A:Map position: 16

Query Match 20.5%; Score 1489.5; DB 2; Length 1612;
Best Local Similarity 28.9%; Pred. No. 2.8e-72;
Matches 402; Conservative 231; Mismatches 516; Indels 243; Gaps 45;

QY 4 PRIHEPMDTTPVKNDPPTFNQCAEGNPTPTIQWFKDRELKTDG---SHRIMLPAGGL 60

Db 29 PRVHEPDLTVSGEPATLNCAGRPPTTIEWYKGSERVETKDKPDRSHMLPSSGL 88

QY 61 FFLKVIHRSR-ESDAGTWEACNEKFGVARSRNATQVAVLRDEFKLPANTRVAQGEVA 119

Db 89 FFLRIVHGRKSRPDSGVICVARNLTGEAVSHNASLEVALRDQFRQNSDWNVAQGEPA 148

QY 120 LMECGAPGSPPEPQISWRKNGQTLNLVGNKRIRI-VDSGNLAIQARQSDGQYCVVKN 178

Db 149 VMEQPPPHGPEPTISWKDGSPLD---DKDERITINGKMLTITRKSDAGKYVCTGN 205

QY 179 VVGTRRESATLKLHVHVRPLIRGQNPATVAGSSVVFQCRIGDPLDPLVWERTASGGM 238

Db 206 MWGERSEVVALVLEKPSFVKRPENLAVTVDDSAEFKCEARGDPVPTVWRXK--DDEL 263

QY 239 PLKFKSWLRSAGRVHLEDRSLKLDVTLDMGEYTCADNAVGGITATGTLVHAPPK 298

Db 264 P-----KSRYEIDHDLKLRKVTAGMGSTCVAMNMGKASASLTVQRPBP 313

QY 299 FVIRPKNVLIGDEVLFPQANGHPRPTLTVSGNSLL---PGYRDMGMEVLTPE 355

Db 314 FVVRKQDVLVAGLRVTFQCEGNGPQALVWRREGSNLLFSYQPPSSSFFSVSTQD 373

QY 356 GRSVLISARFARDSGKVTICNALNAVGVSSSTVSV--DTQFLPPPIEDGQVQPLP 414

RESULT 2
T14160

Query Match 20.0%; Score 1455.5; DB 2; Length 1651;
Best Local Similarity 28.8%; Pred. No. 2e-70;
Matches 406; Conservative 229; Mismatches 525; Indels 251; Gaps 45;

QY 6 PRIEHPHPTVPKNDPPTTNQAGNPPTPIQWFGRDREKLITDG---SHRIMPAGGL 60
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 68 PRIEVPSPDLVSKGEPA TLNCAKGRPPTTIEWKGYGERVETDXDPHSRMHLPSGSL 127
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 61 FFLLVKIHSRR--ESDAQTICCKAEKFQVARSNNATLGVAVLRDEFLEPMATFVAQEVA 119
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 128 FFLRLIVHRGSRRPDGGVICVARNVLKGVASHNASLEVAILRDQPQNSDVNVAVGEPA 187
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 120 LMGSGAPPGSPPEPISWRNGRTQLMLVGNRRIRI-VDDGALIAEQARGSDRGYCVKN 178
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 188 WDCQPPRGEPPEPISMKKDGSPLD---DDRERTIRGKLMITYTKSDAGKYTCVGIN 244
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 179 VVGTRESATAFLKVHVRPPLIRGPQMQLTVGVSSVFQCRIGGDDPLVLRRTASGGNW 238
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 245 WHGERSRYADVTLPERSVFRSPLNLAVDOSAEFCEARGDPTFGWM---DDEGL 302
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 239 PLAKFSWLRSASGRVILVDRSLKLDVLTEDMGYTCEADNAVGGITATGILVHAPPK 298
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 303 P-----KSTELRIDHTLKRAVTAAGDSQTCVADNVMGAKEASTLVIQGPH 352
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 299 FVIRPKQMLVNIQDEYFLBQAGNHPRPTLWVSNGSSL---PGYRGMEVTLTPE 355
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 353 FVVPRQDYVALGRTFTPCBATGNPPALFWRBSQNLLFSQPOSSSRFSVSQTGD 412
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 356 GRSVLISARFAEDSGKVTCNALNVGSSVSTRVSV-DTQFLPPPILBQPVNQTL 414
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 413 ---LTVINVSGSDWTY-CQLWVASGIIIRALETVDVIADRPPIRGPFVNYTA 467
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 415 VKSVILVPCRLTGTPVQVSWYLGIIDVQEHERRNLSDAGALTISLQRHEDBGLTYC 474
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 468 VDGTLTLCVAGTSPPVTLWRKDGVLSTQDSRIQL-ESGVQLR-TAKLGDTGYTC 525
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 475 VASNNGKSSMSGTGLD---TPPNINKIFRAPELTSPPGPKQMWKEGEN 525
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 526 TASTPSGEATHSATIEQVGFVYPVPPRPDNL-----IPASKSPEVDVSNK 575
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 526 SVTLWSRNSVGGSSLVGYTEMKGNETDGAWAGTVRNTFTPTGLLGNVNFYL 585
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 576 TVLLAQ-NPLNLSGATPTSYIEAFSHASGSQMVADVNTETFAIKGLKPNAITLFL 634
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 586 RAENSHELGLSPFMSEPI-TVGYRFNGLDISEARALSLSGDV-ELSNASVYSTMSK 643
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 635 RANAYGISDFDSQISDPVTDQVPTPYGVDRHQVREL-GNVLLHNLFNTLSSSSVE 692
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 644 LTWQI-IMGKVBEGFYVQAQLPNIPLNVPAPTNTNPLMGLSTSSASASAASLIST 702
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 693 VHVTVQQSQSTIQGYKILR----- 712
|::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 703 KPNIAAAGRKGEBTGQGSGGAPPLNTKYRMLTLINGGASSCTITGLVQTLTFYFFIV 762
|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 713 ----PSGASGESBLVFEVFP--TK-----NSVILDLAGYNVTEKAP 753
|::||::||::||::||::||::||::||::||::||::||::||::||::||

31 PVIIIEHPITVWSRGSAPFLNGCAKPS-TAKITWYKQGPVITNKQVSHRIVLDTGSL 89

QY 61 FLFLKTHSR--RESDATGYWCAKNEFVASSRNATQVAVLREDFREPAMTRVAGQY 118

Db 90 FLFLKWSNGKXSDSDAGATCYVSSAWEHGEVKSNEGSIALALMEDFPVAPRPTVQALGGE 149

QY 119 ALMECAFGSPSPQIWSWKNQSTLNVGKRIKRIYVDGNGALQAEASQSDGRYCVYKM 178

Db 150 AVELECSPPRGFPFVWVSKKDKELRIQDMPTRIHSDGNLIDPVDVSDSGTQVCVNM 209

QY 179 VVGTRESATATLKVHVRPFLIRGQNCQATVGVSSVFGCRIGGDDLPVLMHRTASGNN 238

Db 210 MWGEVRSAPLARVVEKAPFEQPKDMTVDGVAAVLPCRVGDPQGPQITWKR--KNEPM 2

QY 239 PLKRFWSLWSASGVEVLED-BSLKLVDVLEDMGEYCEADNAVGGITATGILLVAPP 297

Db 268 PVT-----RATIAKDNRIELRVQPSDBEYTCYARNPAGTLEASHLVQAPP 317

QY 298 KFYVIRKQVLVEIDQVELEQCAANGHPPTLVWSVGNSSLLPGV--RDGMEVLTLP 355

Db 318 SFQTKPADQSPVAGGTATBCTLVQGPSPATFVSKGEQQLLFFSTVSADGKTVPST-- 375

QY 356 GRSVLISARFAEDSKVYTCMLNAVVSUSSTRVVSDTF-----ELPPIITQEGPV 409

Db 376 --GLTILEEVQVDEGATV-CAQMSAAGSSIAKALVITHTAVGTNPAKPPPTTBHQJ 432

QY 410 NQTLPLKSVKVLPCRTIGFPPVQVSWLQDIPIDWGEHRNLSADAGALISQLQREHDE 469

Db 433 NQTLVGGSSALVLCQASGPTPGISWLRDGLDITD-SRIQSHTSGSLHIAADK-POT 90

QY 470 GLTVCASNRMKSSWGLRLDTPINPKIAFFAPELSTPGPGKPKVMEKGENSVTL 529

Db 491 GYTCIAKNEGSESTWASLIVEDYTS-NAQVYRMPDPSFSPSTOPTIIVNVTDEVEL 549

QY 530 SWTRSNGVGGSLGVIEMFKNKNETDGVANGVTQVMTTPTQTLGLPCWNFTLTAEN 589

Db 550 HNNAPSTSGAGPITLIQTSYDGLQGTWENIPDYAVSTETRIKGLKPSHTMFIYAE 609

QY 590 SHGLSLPSPMSEPITVGTIRFNSGL-----DLSARASLLSGDVLSNASVDSMSK 643

Db 610 EXGISTPVSYSALVTISKPAQVALSDKNKDMAIAEKLRLTSLQLLKEVYITNSTATE 669

QY 644 LTQWLN-KTIVBGFYVYARQPNPFINNPPVTSNTNPLLGSTSTASASASASALIST 702

Db 670 LEWKARKLEEIDGTYIKWR-----GPRPTNDQVNVYVSPS----- 706

QY 703 KPNIAAAGKRDGENTGGGAPLNTTKYRMLILNGGAGSSCTITGLQYITLFFTFPV 762

Db 707 -----TENTVSNLMPFTNFFTFPV 727

QY 763 FYK--SVBGSFNSRIARTLEVDSPSAPYMEALLNNSAFVKKWAPELKDRGLVNL 819

Db 728 YBSGVSHIGAPSNHMLTAAAPPPLPEDVRIAMLLTTLWISWAKPAQGIINGLKG 787

QY 820 YHIVRGIDTAHNSRILNTYIDAASPTLVIALTEGMYTVGVAGNAWAGVPT-CV 877

Db 788 PQIVIG-QAPNNK--NITTNERAASYTLPHLVTGHTIKIRVAARSGNSVBSGTS 842

QY 878 PATLRDLPIKRL-----DPFIN--QRQDVNDVLTQPFWILLGAILAVMLSGFAM 928

Db 843 EVNMNQDTLEKHAQQENESFYLGLNKSVF-----VIVIVAILIPVVIYIAT 894

QY 929 FVK-----RKHMQQSALNMRGNHSTVLKMPSS-----LSARNGNYWL 969

Db 895 YWNSRNSGDGRSFRIKINGSVH-MASNNMLVQVANGPNQPNMTAGMTNNRNQAL 953

QY 970 DSST-----GCKVWRPSPGDSLEMQKDHIADYAVCGAPSPAGGTSGG 1016

Db 954 YSLTPNAQGFENKDDYSGMTMRGP-----SEHHYTHAQLTGGPN----- 994

QY 1017 SGGASGASGAGDIGHGHSERNQRYGVEXSIPNTDYAEVSSFGKAPSEYGRHGNASPA 1075

Db 995 -----AMSTF-----YGNQVHDPS 1009

SUBMIT 3
2405
sax-3 protein - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 03-Dec-1999 #ENGLAVE_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42405
R:Zallen, J.A.; Tl, B.A.; Bergmann, C.I.
Cell 92, 217-227, 1998
A:Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple axon
A:Reference number: Z22160; MUID:98117250
A:Accession: T42405
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <AL>
A:Cross-references: ENGL:AF041053; NID:g2804779; PIDN:AAC38848.1; PID:g2804780
C:Genetics:
A:Note: sax-3
C:Function:
A:Description: sax-3 function is required at the time of axon guidance

Query Match 18.7%; Score 1361.5; DB 2; Length 1273;
Best Local Similarity 28.4%; Pred. No. 1.6e-65;
Matches 357; Conservative 202; Mismatches 437; Indels 259; Gaps 37;

Qy 4 PRIIEHPMDITVPKNDPFTFNCQAEGNPTPTIQWFKDREL---KTDTGSHRIMLPAGGL 60
| | | | : | : | | | : : | | : | | : : | | | | : | |

Qy 1077 PYATSSILSPHQ-----QQQQQPRYQRPVPGYGLQRMPH-----HYQQQQHQQQQA 1125
 Db 1010 PYATTTVLNSNQPAWLNDKMLRAPAMPYFVP-----PEPPARYADHTAGRRSSRSA 1063

Qy 1126 QQ-----THQQHQALQHQQLPPSNI-YQQMSTTSEIYPTNGPSRVSYSQ 1171
 Db 1064 SDGRGTNLGLLHRTSGSRQSDSPPHYDVTQVGLHSSD-----GTGSSKERTGER 1113

RESULT 4
 T14316
 rig-1 protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14316
 R:Yuan, S.S.F.; Cox, L.A.; Dasika, G.K.; Lee, E.T.H.P.
 submitted to the EMBL Data Library, April 1998
 A:Reference number: Z17975
 A:Accession: T14316
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 Molecule type: mRNA
 Residues: 1-1344 <TAB>
 A:Cross-references: EMBL:AF060570; NID:g4206385; PID:g4206386; PIDN:AAD11628.1

Query Match 17.34; Score 1256; Db 2; Length 1344;
 Best Local Similarity 27.24; Pred. No. 8.3e-60;
 Matches 376; Conservative 194; Mismatches 462; Indels 356; Gaps 48;

Qy 4 PRIEHMOYTVFANDPFFNQAEQNPPTIQWKGRELK---DTGSHRMLPAGUL 60
 Db 42 PRIVGPPVPLVSGEPAFLPCRASGRPRPILKWTYKNGARVATAREPRHRLLLPSGAL 101

Qy 61 FFLKVIHRSR-ESDQGTCEAKNEFVARSNAITQAVLDRLEFLEPANTFVAGSVA 119
 Db 102 FFFPRIVGHSRSPDQGYTCVARNILGAASRANSLSVALRDRDQFQSGPQNVVAVGEPA 161

Qy 120 LMGCGAPGSPERQISWRKNGQTLNLVGNKIRIVDGGMLAKQEARSDGRTYCVVKNY 179
 Db 162 YMGCVPPGHPLEPLVWKKG-KIKLKEBGRITTAGKLMHSHTFYSGAGMTCVASN 219

Qy 180 YGTRSEATAFLVKHYVRFLIRGQNTAVGSSVVFQCRIGGDPDLVLRRTASGGM 239
 Db 220 AGERESGAELVLRPSFLRRPQNVVLADAPVNLCEVQGDQPKLHWRK-DGGLP 277

Qy 240 LRKFSWLHSAQSRVHLEDRSLKDDVLEDMGSEITCEADNAVGITATGLTVRAPPK 299
 Db 278 -----AGRIEIRSDSLWIDQVSEDEGTYTCVANSVGRAASGLSVYPPQF 327

Qy 300 VIRPKQWLEIDGDEVLEFCQANGHPRLPLVNSVGNSSLLLPYDRGMREVTLTPGRSV 359
 Db 328 VTKPQNTVAPGANVSFCQETKGNPPPAIFPKQKESQVLFLFSQ-----SLQMGKLL 380

Qy 360 -----LSIARFAREDGSKVTCNALNAVGVSSRTVSV-DTQFLPPPIIEQGPVMT 412
 Db 381 VSPRGQLNTEVILGOGGTY-CQAVSVAGSLAKALLEIKGASIDGLPIILQCPANT 439

Qy 413 LPVKSIVLPCRLTGPVQVSN-----YLDGIPIDVQHERHNSDAGALTISDLQHE 467
 Db 440 LVLGSSVWLPCRVLGNPQNIQWKKDERWLQ-----DOSGPDNLMDNGHILASIQ-EM 492

Qy 468 DGLNITCVANSNRKNGSSNGYLRLLT-----PTNPINIKFRAPELSTYPPGPK 516
 Db 493 DMGPITSCVANSIGERATWNSLWKLEDQNGASGPATGSPNP-----PGPPSQ 539

Qy 517 PQWVEKGSVSLTWSRNSKVGSSVLGYVLEMGKNETDGVAVGTQVNTFTITGELL 576
 Db 540 PIVTEYANISLTILN-KPNQSGATATSYLEASQAGANTWRTVADVGQLETITISGLQ 598

Qy 577 PGVNTYILFRAENSHGLSLPSMSEPITVGTITFNSGLDSEARSL-----LSDVD 628
 Db 599 PNTYITLVFRVAGWGLSEPSVSEPVQT-----QDSLSRPAEDPWKQKQGLAEVA 650

Qy 629 VELSNASVDSSTMKLTQWING--KYVEGFYYAR-----QLNPPIVNNAP 674
 Db 651 VRMQEPTVLGPTLQVSN-TVDGPVQLQGFQFVSNRIAGLDQGSWMLDLQSP-HKQST 707

Qy 675 VTSNTPLGLSTSTASASASASALISTKPNIAAGKRGDGTNGSGGGAFTPLNTKMYL 734
 Db 708 VLRLGFP-----GAQIQKQVQ-----GGQGLGASFPFVTR---- 739

Qy 735 TILNGGAGSSCTITGLVQYTLFEFFIVPFYSVGEKPSNSIARTLEDVPEAPYMEAL 794
 Db 740 -----SIP-----EAPSPQPGVAVA 756

Qy 795 L-LNSSAVFLWAKAPELADRHGVLLNHYVHVGIDTAHNFSRILTNVTIDAASPLVLA 852
 Db 757 LGGDRNSVTVSNEPPLPSQRMGVITEYQICWLQ-----NESRPHLNRSAQWASVTVS 811

Qy 853 NLTEGVMTYTVAGNNGAGGPGYCVATRLRD-PTTKRLDPFNG--RDHVNVLVQPFW 909
 Db 812 GLLPQITRAIVANATSAGVGVASAPVLVQLPPPPAEPQGVPSGLAERLAKLPKAP 871

Qy 910 IILLGLAILAWLSPGAMVYFKRMMKMSALNTMRGNTSDVLMPSLSARNGWT-- 967
 Db 872 LAGSSAACALLGFCALYRQK-----QRKLS-----HYTASFTATVPSPHSELG 923

Qy 968 -----WLDSTGWNWRSPSGDSELMQKMDIADYAPWGAPGSPAGGOT 1012
 Db 924 SSSRPWGLGPAATYWLADS-----WPHPPRSFAQEP-----GSCCPNPPODORY 972

Qy 1013 SSGG-----SGGAGSGAG-----GDDHGGHGSERNGYQVTSNIPTDYA 1055
 Db 973 NEAGISLTLAQARGANASGSPVSTIDPVEELTFHGG-----FPQHSQSGSPWS 1026

Qy 1056 EVSSFGKAPSEY-----GRHG-----NASAPYATSSILSPHQQQQ 1092
 Db 1027 QT-----APPNSGEGSARGQGLLKGVPQWLSLVEALPPPPSCELSCPEGE 1061

Qy 1093 QQ-----PRYQRPVPGYGLQRMPHPTQQQH 1120
 Db 1062 LKGSSELEWCPVPEKSHLVSSSGSNCAVAPAPADPTSPSTSG-----1127

Qy 1121 QQQQAQQTHQHQALQHQQLPPSNI-YQQMSTTSEIYPTNGPSRVSYSQITPKK 1178
 Db 1128 QQSTALITLPSPPDPQ-----PTDIPHLHQMRRVPL-----GPSPSLVSQPALSGH 1177

Qy 1179 QRHRIHTENKLSNCE-----TYEAPG-----AKQSPISQSFASVRQ-----1217
 Db 1178 GRPWGLAGPVLSTHASPSPVSTASAPGRTQVTGEMTPLAGHRARIKKPKALPYR 1237

Qy 1218 -----QLPP 1221
 Db 1238 REHSPGDLPP 1247

RESULT 5
 T29549
 hypothetical protein ZK377.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T29549
 R:Nhan, M.; Hawkins, J.
 submitted to the EMBL Data Library, February 1997
 A:Description: The sequence of C. elegans cosmid ZK377.
 A:Reference number: Z20639
 A:Accession: T29549
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 Molecule type: DNA
 A:Residues: 1-423 <NAB>
 A:Cross-references: EMBL:U88183; PIDN:AA852658.1; GSPDB:GN000028; CBSP:ZK377.3
 A:Experimental source: strain Bristol N2; clone ZK377
 C:Genetics:
 A:Gene: CBSP:ZK377.3
 A:Map position: X

A:Cross-references: EMBL:U11021; NTD:q562122; NIDN:AAA673607 1; NID:q562122

Qy 452 LSDAGAL--TISDLQRHEDEGLYTCVASNRNGKSSWSGYLRDLTPNPNIKFFRAPELST 509

Db 479 TSQPSISLQITVGNLTPEETIN-PRVATYENWGPGE-----SQSEKVVYTOPELQV 527
 Qy 510 TPQPPKQPMQVKEGNSVLTWTRSNKVGSSLVGVYI---EMFGNKETDQWVAGTVRQ 566
 Db 528 -PGVVENLQVSTAPTSVLISNDPPAYANG-PVGGRLCACTSGRQNG-----IYD 579
 Qy 567 NTYTPQTGLLPGVNYVFLRAENSHGLSLPSMSEPITVGTIRFNSGLDSEARASLSG 626
 Db 580 GIVTRLEGRLKPTSTSRVLAENRYG---PGVSEHRTVYV-----LSDVPSAMPQN 628
 Qy 627 DVVLESNASVDSMTSLKLTW-----QIINGKYVEGTV----- 659
 Db 629 VLSLEVN---SRISIKVWMLPPPGTQNG-FITGKIRHRKTTIRGELETLNPNLWYL 682
 Qy 660 -----YARQLPNPIVNNPAP-----VTSNT--NPLIGSTSTASASASALIST-- 702
 Db 683 FTGLEKGSQYSFQVAMTVNGTGPSSDWTATETPENDLOESQVPPQPSLVRPRLTISII 742
 Qy 703 -----KPNIAAGKRDG-----ET-----NQS 719
 Db 743 MSWTPPLNPNIVYGGIIGYGVGSPATZTVNDQKQRTYSIENLEPSSHYISLAKFNA 802
 Qy 720 GGGAP-----TPL-----NTXFM 733
 Db 803 GEGVPEVIESATTSQTVPMDSMTPMLPPVQVQWALHDAVRVSWADNSVLKQKTEVERP 862
 Qy 734 LTI--LNGGAS-----SCTILKQVYLVKEFIVPYTKSVGEKPSN---SRI 776
 Db 863 YTIWRTSTYSASSKYSADTTLSTSHVYGLKPNVTFESVW---VTGRSSSTWSMTAH 918
 Qy 777 ARTLEDVPSFQVGMHALLNS--AVLKAWAPELDRHGLVLANVYVIGDIAHNPS 834
 Db 919 ATTYETAPTSAKQLVITREKPRANIVSQWPP--IIRANGKIIF--ILFTYLDKMLQLD 975
 Qy 835 R-LTWITVLDASQVLAANLEGVYVGAAGNAGVGCVPATLRLDPIIKRLDFF 893
 Db 976 DMWITVITGURLTEILDLLN--DTATYFRQANNAAGLPSLEPTFR-TPKVEPDRKM 1032
 Qy 894 INQDRVND-----VLTPQ-----WFILIGAIL 917
 Db 1033 ANDGSGGGGGVSDVNLDRSSLNBPQMHGPPVTPQKNMNLVIVVTYGAIT- 1091
 Qy 918 AVLMLSGFAMVYKRNKMMQKALNTRGHN--TSQVLKMPSLRANMGVTLWDSGG 975
 Db 1092 TILWVIVAVICTRQSSAQKAKRATHSAGKKGSKDLRPPDL-----WIHEEME 1143
 Qy 976 K-WVPSPGGDS--LEMQKRIADYAVPCGAPSGAGGTSSGSGGASGASGGDIH 1031
 Db 1144 MNKTEKPS-GSDTGDSGPRQSCQDITFVSHQSQESGLKSKSTSHSG-----PDADV 1196
 Qy 1032 GGGGS-ERN--QRTYVGEYSNIPDYAEVSGFKAPEYGRHGNASAPATYSS---IL 1084
 Db 1197 SNTISLERTLAARATRAKMLPMD-----SQSSNPVPSAIPVPLESAQPGIL 1248
 Qy 1085 S-----PHQOQQQQQPRYQQRVPV-----GYGLRPMHPYQQQHQOQQAQTR 1130
 Db 1249 PSPQTGTPH-----PQFLRPVFPPLTVDRGFTSRVTEVPSQSSVLSHPQEPH- 1300
 Qy 1131 QHQALQHQQLPPSNYQOMSTTSEIYPNTGPSRSY 1167
 Db 1301 -----STSEDAPSRTIPTACY 1316

RESULT 9

743027

neural cell adhesion molecule L1 - goldfish

N:Alternate names: E587 antigen

C:Species: Carassius auratus (goldfish)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000

C:Accession: 743027

R:Giordano, S.; Laessle, U.; Lottspeich, F.; Stuermer, C.A.O.

submitted to the EMBL Data Library, April 1996

A:Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule

A:Reference number: 222294

A:Accession: 743027

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1232 <GIO>

A:Cross-references: EMBL:U5211; NID:g1305526; PID:g1305527; PIDW:AAA99159.1

C:Superfamily: neural cell adhesion molecule L1: fibronectin type III repeat homology

C:Keywords: cell adhesion; membrane protein

Query Match 8.24; Score 593; DB 2; Length 1232;
 Best Local Similarity 23.24; Pred. No. 3.7e-24;
 Matches 296; Conservative 146; Mismatches 495; Indels 340; Gaps 51:

Qy 4 PRIEHPMDTTPKNDPPTTFQCAQGNPTTPIQWFKGRELATQDTSRIMLPAAGLFFL 63
 Db 15 PTITVQVPSHTAFSLDDVLCAEASGDPAPSRFWKVGKKEFK-----RELLSSGTL-- 65
 Qy 64 KVHSRESD--AGTYWCAKNEFQVAKSNAL--QVAVLRDEFRELPANTR-VAQGE 117
 Db 66 -TAEDEKELHPIQSQRYCYVM-LCTAVSLQALITEPTIPLAKERK---QRTASFEQD 120
 Qy 118 VALMECGAPRSPPEQISWRKNGQLMLVGNKRRI-VG--GNLAIQBARQSDGGY 172
 Db 121 SAYLKNPKPSVTPKIH-MDMWRHPIPLNEVTTSLDGLYFANLLWDSRED---Y 175
 Qy 173 QC-----VYNNVGTRESATAFLKVVHRP---FLIRGPQW-----TAVIGSVYF 215
 Db 176 TCNARIHNASVILPKER---ISISVTPSNVLKNRKPLQKAGSHSSVYLKQVTL 231
 Qy 216 QCRIGDPLQVLAIRTESAGGNMPLKRFVSLHSGASGRVYVLE-DSGLKLDVLEDMGEY 274
 Db 232 ECIPEGLPTFQVQWEMDS---PL-----SPARVWLAKRWKQVTSSEADGGE 279
 Qy 275 TCADANVAGISTAGILVYAPKFEVIRKQVLGVDEVLFCQANGHPPTPLVSWEG 334
 Db 280 TCTAQNSGQSKVHTVATYEAAPYTRVLAPEVRLDQAGGIPPTNITWGM-- 337
 Qy 335 NSSLLPPTROGMEVTLTPGRSVLSIARFARDESCKVVTCTNALNVGSSVSRVTSVD 394
 Db 338 NGAPLAGTDPDRKRVYS--SCTLILTDVQIS---DTATYVETANKHKLINLTHRVH- 390
 Qy 395 TQPELPPITPDGPNWQTLKVASIVLPCRTLTPQVPSYVLDGIDPIDVQEHKRNLS 454
 Db 391 --VELPQILFDOLKATEBGQVLLQCRITFGSPQPKVDQITNSGPAIANKAKQSQSD 448
 Qy 455 AGALTSISDLQREHDEGLYTCVANSRNGKSS----- 484
 Db 449 -GNLQISDVS-DEDSMTYCTSSPNNMSIAELVNLNRTKIVDPQDLRVLAGDQVLQ 506
 Qy 485 -----NSGYLRDTPTNIKIFFRAP-----ELST-- 509
 Db 507 RITVDMLKQITQWKKDKHKTSSANDKTSFSGSLKITDQMEDSGISCEITKL 566
 Qy 510 -----TPGPKQPMQVKEGNSVLTWTRSNKVGSSLVGVYIMFGK-NET 555
 Db 567 DVSATSGSVYLDKPGPSHLSLSCKKRSVLTWNMPGAE-NNSPISEYVIERKQKQPC 625
 Qy 556 DG-WAVGTVRQNTITFTQGLPGVNYVFLIRANSHGLSLPSMSEPITVGTIRFNSGL 614
 Db 626 KGHMEYRVPQDITHEILHQPITYSTHFVRGTVNGHSEPSSESTST----- 676
 Qy 615 DLSEARASLLSGDVVELSNASVVDSTSMKLTWQIINGKYVEG-----PYTARQ----- 663
 Db 677 --PAAPKQPNFNTYSVS-----TDSNLSVITQELBQRFQKPGFKYIKITWQEQSDHNM 730
 Qy 664 ---LNPV--IVNNPAP-----VTSNTNPLIGSTSTASASASALIST-- 693
 Db 731 ESSASNPPTFVEGCTTFFQIKVQVWNLGAEPEADRGISYSEDLPLAEPASVAVSEL 790
 Qy 694 ASASALSTKEP--NIAAG-----KRGDTWNGSGGAGPTPLNTYMTLNTLNGGA 742

```

RESULT 11
T30532
neural cell adhesion molecule L1 homolog - Fugu rubripes
C:Species: Fugu rubripes
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T30532 ;
R:Riboldi Tunicliffe, G.R.; Platzer, M.; Nyakatura, G.; Elgar, G.S.; Brenner, S.; Ro
submitted to the EMBL Data Library, September 1997
A:Description: Analysis of the genomic loci of Fugu rubripes homologs of the human di
A:Reference number: J20848
A:Accession: T30532
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1277 <R1B>
A:Cross-references: EMBL:AF026198; NID:a3098263; PID:a3098264; PIDN:AAC1580.1

```

A: Note: L1-CAM

Db 1009 MARTAAGKG 1017

Db 897 TNLRPYCHYDLAISAFNSKGEGPLSEKTSFMTPEGVPG-PPMSMQMTSPSESEITLHWTP 955

J. Polym. Sci. Part A: Polym. Chem. **37**, 109–116 (1999)

Qy 252 RVHVLDSRLSKLDVTDLEDMEGYTCADNAVGGITATGTLVHA--PKFVIRPNKLQVEI 310
 Db 367 RIEV-SGGELRFSKLVLDSSGMYQCAENKHTVYASAEITVQALAPDFRLNPKRLIPA 425
 Qy 311 --GDEVLFEQAGHPRTLYSVNBSGNSLLPQYRDMGEVITLPEGRSVLSIARFAE 368
 Db 426 ARSGVYIIPQAPRAKATVMT--KGTGLNSSR-----VTIADGTLILQ--NHSKS 476
 Qy 369 DSGKVTCTNALNAVGSVSSRTVVSVDIQFPELLPIEQGVNQLPWSKSVILPCLRTGT 428
 Db 477 DGGK-YTCAENFMKXANSTGLSVDRAKT-----ITLAPSSADINVENMLTQCHASHD 530
 Qy 429 PVQVQ--SWYLDGPIIDVQHE-----RRWLSDA-GALTISDLQ-RHEDEGLTCAVSNRN 480
 Db 531 PYMDLTFVNSLDPPIDLDKSEGHYRASGVKAVGDLAINQAQKH--SGRTTCAQTVV 588
 Qy 481 GKSNSWGLRLDPTNPNIKFTRAPELSTYPPGPKQPMVEKGENSVLWTSRKNKVGGS 540
 Db 589 DSTSESATLTVRG-----PGPPGVVVVDIGDQTLVLSWRGDF-NHS 631
 Qy 541 SLVGVEIE--MFGNKTDGKVAVGTRVN-----TPTOTGLLPGVNTFLIRAENSH 591
 Db 632 PIARTYSIEARTLSNK-----WKQMRTPVNIENGAETAQVNLIPWDTYFVRVLSNML 686
 Qy 592 GLSLPMSSEPIITVGTTFNSGLDSEARASLDGVDVLSLNSVDSSTSMKLWTQIING 651
 Db 687 GVGSEPS-----PGPPGVVVVDIGDQTLVLSWRGDF-NHS 692
 Qy 652 KYVEGTYVYARLPNPNVNPAPVTSNTNLLGSTSTASASASALISTK--PNIAA 708
 Db 693 -----LP-----SSKRTKEAAPVAP 709
 Qy 709 AGKRGHTYNGSGGAP-----TPIMLYKRLMILNGG-----ASSCTITGLV 751
 Db 710 SSL-----GGGGAPNELLINWPTLIDYQ-----WDGFGYILSPKKGTQGWKATRV 758
 Qy 752 -----QITLYEFFIVPFYKSVBEGKPSNRIARLNDVSEAPYGEALL 795
 Db 759 PHARSHTVYRNESIGFTTPEFKIKATNRKGSGPSLTAIYSAEPEYAPVRYAKA 818
 Qy 796 LNSAVFLWKAPELKDREHVLNHYH-IVRGDIHANFSRLINVTIDASPTVLANL 854
 Db 819 VLSSEVDVSNPEVQGDMTGVLLGTETIRYWDGKEAARDVRIAGLTVSAHTV-----GL 874
 Qy 855 TEGMVTGVVAGMNAGVGPIC-----VPATLRDPTIKRLDPTFNG 896
 Db 875 NPMKTVHSVATYNRAGAGPSPSTNITTTKPPRPGLNLSWTLGSTVYTIKNDVPAQ 934
 Qy 897 Rd 898
 Db 935 Rd 936

RESULT 13
 A43425

Bravo/Nr-CAM cell adhesion molecule L1 homolog - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C:Accession: A43425
 R:Kayman, J.F.; Roman, J.M.; de la Rosa, E.J.; Schwarz, U.; Dreyer, W.J.
 J. Cell Biol. 118, 1259-1270, 1992
 A>Title: Bravo/Nr-CAM is closely related to the cell adhesion molecules L1 and Ng-CAM at
 A:Reference number: A43425; MUID:92381110
 A:Accession: A43425
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid; protein
 A:Residues: 1-1259 <KAI>
 A:Experimental source: cerebellum
 A>Note: sequence extracted from NCBI backbone (NCBI:112026)
 C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homolog; 4
 F:237-294/Domain: immunoglobulin homolog <IMD>

Query Match 7.8%; Score 566; DB 2; Length 1259;
 Best Local Similarity 21.0%; Pred. No. 1.1e-22;
 Matches 279; Conservative 180; Mismatches 514; Indels 354; Gaps 51;

Qy 8 EHPMDVTVMNDPFTTQCAEQNPTPTIQWKGDELKTLTGSHRMLPAGGLFVLAHV 67
 Db 22 QSPKDIIVDPENITVQCEAKGPPSPWNRGTHTFDINDQAQVMTKPNGLVNLWM 81
 Qy 68 S-RRESADTYWCAENKNEFVARSNATL--QVAVLDEFELEPANTVRAQVAGEWMEG 124
 Db 82 GVAKAEVGYQICTARNEPGLAISNVIYKXSPILWTKKLEP NHRV--BGDSLWLNCR 139
 Qy 125 APRGSEPEIISWRNQTNLNLWKNRIRIVDGNLAIEQARQSDGR--TQCVW--NVV 180
 Db 140 PPVGLPPPIIIPWMDNA-FQRLPQSEVSGNLGDLVFSNV-QPDTREDVICTARFNETI 197
 Qy 181 GTRESATFLAV-HYRPFLLRGP-----QNOQTVAGSVVYQCRIGGDLPLVLMR 230
 Db 198 QQKQKQISVYFVSTKPTTERPVLITMGSTSNKVELGNVLLLECIAGLTPVIVRWI 257
 Qy 231 RTASGNNPLAKFSWMLASAGRVHVLDSKLDVLEDMGYTCADNAVGGITATGI 290
 Db 258 K--EGSELPAKRTFFEN-----KTKLTIIDSEADSNYKCTARNTLSTGHVYS 306
 Qy 291 LTVHAPPFYVRPNKMLVEIGDEVLEPCQANGPRPTLYSVNBSGNSLLPQYRDMGEV 350
 Db 307 VTVAAPTYWTAFAKMLVLSGEDGLICRANGPKFISWNLGTVPIALAP--EDPSKV 364
 Qy 351 TILPEGRSVLSIARFAEDSGKVTCTNALNAVGSVSSRTVVSVDIQFPELLPIEQGVN 410
 Db 365 -----DGTIISA--VQERSSAYQCSNNEYGYLLANFVNLAE-----PPRLT--PAN 413
 Qy 411 QITLV--KSTVLEPCLRTGTPVQVSWYLDGPIIDVQHEHRLNSDAGALISDLQHEH 468
 Db 414 KLIQVADSPALIDCATGSPKEPEW--RGVKSILAGNEVVFMDGTETLIPVAG--DS 471
 Qy 469 BGLTCTVANSNKGSSNGYLAALDPT-----NPNIKFFRA-----PEL---- 507
 Db 472 TGTTCTVANKRLGTQNEVLEQDVTMIMQIQVYKVSQAQSPFCVIXHDPITLPTV 531
 Qy 508 -----STY----- 510
 Db 532 IWLKDNELPDERFLRGKDNLTIMNVDKDDGTTTCVNTVLOSVASAVLTVAAAPT 591
 Qy 511 -----PGPGPKQPMVEKGENSVLWTSRKNKVGSSLLGVTVIEM-PGKNETDGVAVG 562
 Db 592 PAITATVAPNPLDLELTGLQLERSIELSVWPEG--NNSPINFVITVEDGHEPGVHYQT 650
 Qy 563 TRVQNTTITQTLGPNWYFFLIRAEENSHGLSPSPSEPIITVGTTFNSGLDSEARAS 622
 Db 651 EYVSGTITVQLKLSPTVNSFRIANKEIGRSQSPSPS-----QITLKSAPDE---- 700
 Qy 623 LLSGDVWLSNASVDS--TSMKLWQIINGKYVEG-----PTVYARLPNPNVNPAPV 676
 Db 701 -----NPSMVGIGSEPDNLVITWESLKGFGNSGPGQVQVSWRQKDDVDSTVVA 753
 Qy 677 SNTNPLLGSTSTASASASAL--ISTKPNIAAKRGDE-----TN 717
 Db 754 NVSKITVSGTIPVTEIKVQALNDLGTAPPEPSVHSGEDLPKPAQNVQVHINSTL 813
 Qy 718 QSGGGAPTL-----NTRYKMLTILGGGASSCITITGLV 751
 Db 814 AKVHNVPPVLESVRHGLQYKYVYVQVSLSRSSKRVKVKLLT--RGKNTFGMLGLE 871
 Qy 752 QITLYEFFIVPFYKSVBEGKPSNRIARLNDVSEAPYGEALLNSAVFLWKAPEL 811
 Db 872 PISTKMLVHYVNGKGBGSPADVKETPEGVPS-PPFLKINPTLSDILWESGP--T 928
 Qy 812 DRBGVLLNHYVIRGDIHANFSRLINVTIDASPTVLANLDEGVMY-----TWG- 863
 Db 929 HPNGVSYLILFQPINNHELGP-LVETIPANESSLLNMLNSTRYKFTNAQTSVGS 987
 Qy 864 -----VAGNAGVGPICVPATLRDPTIKRL-----DPFIN-----QRQVW 901

Db 988 GSQITEAEVITDIEAGILRPVAGAG-KVQPLPRIRNVTAAETIANISWEYEGPDHAN 1046
 Qy 902 ----- 901
 Db 1047 FYVEYGVAGSKGDKKKEIVNGSRSPFLKGLTPTGATYKVRVAGBLSGFRASSELFTETGP 1106
 Qy 902 -----DVLTGQFILLGAILAVLMLSGFAMVFKRHHMMKQSLANTMGRNCTSDVL 954
 Db 1107 AMASRDVDTAQWFIQMLCAV-ALLILILLVITFCRNN-----KGG----- 1147
 Qy 955 KMSLSARNGNMGWLDSTGGMWRPSPGSDLMKQDIADTAPVCGAPSGAGGTSS 1014
 Db 1148 KYPVKKEDAHK--DPRIQMKEDDTFGFETRSLESD-AEDHKPLAKGSRTPSDTVKK 1203
 Qy 1015 GSGGGA-----GSGAGGGDDINGGSGERNMQRYVGEYSWIPDYAEVSSGKAPSEGRH 1070
 Db 1204 EDSDDSLVDYGEVNGQFNEDGS-----FIGQYS-----GKKEKPAE-GNE 1244
 Qy 1071 GNASAP 1077
 Db 1245 SREASP 1251

RESULT 14 S26180

neurofascin - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jan-2000
 C:Accession: S26180; Wolff, J.M.; Frank, R.; Rathjen, F.G.
 J. Cell Biol. 116, 149-161, 1992
 A:Title: Structure of the axonal surface recognition molecule neurofascin and its relation to the cell adhesion molecule Li; fibronectin type III repeat homology; L
 A:Reference number: S26180; MUID:92317154
 A:Accession: S26180
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1272 <NDB>
 A:Cross-references: EMBL:X65224; NID:63659; PDB:CAA46330.1; PID:63660
 C:Superfamily: neural cell adhesion molecule Li; fibronectin type III repeat homology; L
 F:279-336/Domain: immunoglobulin homology <IMM>

Query Match 7.84; Score 565.5; DB 2; Length 1272;
 Best Local Similarity: 21.18; Pred. No. 1.2e-22;
 Matches 293; Conservative 160; Mismatches 455; Indels 479; Gaps 55;

Qy 7 IEHPMDITV-----PKNDPTTFNQAGNPTPTIQWKGDR----- 42
 Db 26 IEVPLDSNITQSEPPPTITQTSVKDYIDVDRDNIP-IECEAGNVPVPTFMKNGKFFN 84
 Qy 43 -----ELATGSHRMLPAGGLFPLKLVHSRRSDGATYCEAKNEFGVARSNATL 95
 Db 85 VAKDPKYSKRRSGSLTVIDFEGGG-----RPDYEGEYQCFARDYGTALSKSLH 135
 Qy 96 QVAVLDEFLEPAN-----TRVAGCEVALMECGAPKSGPEQPSVWRKNGQIMLVGNKRI 151
 Db 136 QVS-----RSLPMPKKEVDVIEDEGAPLSLQCNPPPLGPPVIFMSSSME-PIHQDKR 190
 Qy 152 RIVDGG-----NLAIQEARSGDGRGYCVK-----NYVGTRESATFLK 191
 Db 191 SQGQNDLIFSNNMLQDA-QTD-----YSCNARPHETHIQQKRLVLKVKTKPNETSLR 246
 Qy 192 VHRPFLRG-----PQNTAVGSSVVFQRCIGDGLPDLVLRRTASGN 237
 Db 247 NHTIMTSARGVTTETPSMYPYGTSSQWNLKVDLLBCLASGVPAIDWTK-KGGE 304
 Qy 238 MPLKRSNLSASGRVHVL--DRSLKLDVLEDEGYTCRADNAVGGITAGILVHA 295
 Db 305 LPAGTK-----LENFKALRSIWNSEDSGEYFCLASNMGSIHETISVHWKA 353
 Qy 296 PPKFVIRPKNLVIEGDVLEQANGHPPTLKWSEGNSSLLPGYRGRKEVFLPE 355

Db 354 APYWLQEPQNLILAPGEDGLVCRANGNPKSPQNLWNGEP-----IEGSPFPN 402
 Qy 356 GRSVL-----IARFAREDGSKVYCNALNAVGVSSSRVYVSOTVLPFPIIEQGVNQT 412
 Db 403 SREVAGDITVEDTQIGSSAVYQCNASKNHYLIANAFVS-----LDVFFRIL-APRNL 457
 Qy 413 LPV--KSLVLPCTLTGTVPVQSVYLDGIPIDVQEHERRNLSDA-----GALTISD 462
 Db 458 IYQVITNRLRDCDFPGSPITLRFKNG-----CNQMLDGGNTKARENSLEMS- 507
 Qy 463 LQRHEDEGLTYCVASNRNGKSNWSGSLRLDPTNPNIKFFRAPE----- 506
 Db 508 MARKEDQITTCVAINILKGVQVRLVKDPT-----RIVRGPEQVVKRGSMPALHCRV 563
 Qy 507 -----LSTY----- 510
 Db 554 KHDPTLALITVWLADAPALYIGNRMKKEDDGLITVIAEKDQDGTVCSTAVEKLDKAKA 623
 Qy 511 -----PPGPGQPMWKEGNSVULSWTRKMGVSSGLVGYVIMFG 551
 Db 624 YILVLAIPANRLDLPKRPPDRPLESLDLSAERSVKLWIPGCD--NNSPITDIYQFEE 682
 Qy 552 KNEDTG-WVAVGTQVQNTPTQTGLPCGVNTPFLIRASHGLSLPSMSEPIITGIRY 610
 Db 683 DRPQQTWNNSRIPGWNNSALLSLPTVNTQFVIVAVNDGSSLSMPSE-----RTQ 736
 Qy 611 NSGLDLSAASALSGDVELSNASVDSKMLTQIINGKYVEG-----FTV--YAROLF 655
 Db 737 TSG-----ARFELNPTGV-----GQAGTCKNMELITWPLNATQAGQPNLRIYVRRKPD 787
 Qy 666 ----NPIYKNPAPVTSNINPLGSLTSTISASASASALISKFNIAAAGKDGETWQSG 720
 Db 788 RGSWYKTYKAPRHWVNT-PIY-----VPEIKVQAGNDFGAPSEETITIGSG 836
 Qy 721 GSAPTPLNTKIRMLTILNG----- 739
 Db 837 EDYKAPATDVR-IKVLNSTALITWTRVHLDITQQLKETATFWRDSSLLANLWYSK 895
 Qy 740 -----GGASCTITGLWQTLTLEFFIVPFFYKSVGKPSNRITAEVLEVPAGE 792
 Db 896 RQVYSPFGDRNGILVSRLLPYSNKLELVNVTNGDGPSEYKPEPTPEGVSSPFI-LR 954
 Qy 793 ALLNLSAVALKWAPELKDORGLVNLNHYVR-----GIDTAHNSRLITNVIDAA 845
 Db 955 IRQPNLESINLWDHPE--HPNGVLGYNLRYQAFNGSKTGRITVENFSPQWTHVQRT 1012
 Qy 846 SPITVLANLTGGMVTVVGAAGNAGVGYCV-----PATLRDIPITKRLDP 892
 Db 1013 DPI-----SRTFRFLRAQTQGDGEIVESFALLNEATPTPASTLPPPTTELTP 1063
 Qy 893 ----- 892
 Db 1064 AATATITTTATPTTPTPTPTPTPTTTTTTATAASTVASTTTAERAAAAATKQE 1123
 Qy 893 PIHQORHVDVLTQWFIILLGAILAVLMLSGFAMVFKRHHMMKQSLANTMGRNCTSDVL 946
 Db 1124 LATYKNV-DIATQWFIQMLCAI-ALLVILLVITFCRNN-----KGG----- 1147
 Qy 947 GNHTSDVLKMSLSARNGNMGWLDSTGGMWRPSPGSDLMKQDIADTAPVCGAPSGAGGTSS 1014
 Db 1182 KNVEDGSPFYSRLESDEN-----KPLNFSQTLDTITQGESDSDSLVDIT-- 1226
 Qy 1000 VCGAPSGAGGTSSGSGGAGSAGSGDINGGSGERNMQRYVGEYSWIPDYAEVSS 1059
 Db 1227 -----GEGEGEGFNEDGS-----FIGQYT-VKORKEETS- 1254
 Qy 1060 FGAPSE 1066
 Db 1255 -GNESSE 1250

RESULT 15
 I50600

QY 10 PMOTTVPPTTNNQAGNPPITPQWKGREKLTGDSHRMLPAFFGLFKVHSR 69

Db 25 PMOLLVSFGASIVLNCSSCYCTPPKIDWKQDITNLNVSDRRQLLPDGSLLNSVHXS 84

QY 70 -RS DAGTVTCANE--FVGARSNAQLAVAREEFLREPANTFAAGEVALMEACAPR 127

Db 85 HNKPDGEQTQCVNVESLSGSVSTAKLVAGL-PRFTSQPELSVTKGSAINCNEY-N 142

QY 128 GSPEPIISWKNQOTULNLVMNRIRVDGNALAEARQSDGRYCVVMVKVTRESAT 187

Db 143 VDLAFPLFWNEQRDPISL--DWRVFLPSCALLGNATIDGGGYTRCVIESGGTFKYSEE 200

QY 188 AFLKVVHP-----FLIRGPNQATVAGSSVFECRGIGDDLPVLWRRITASGMMPLR 241

Db 201 AEKLILPDEPEQSULTVTPQSSLTUFGQNAPVCVAGGPPTPVYW-TKNSEEL--- 255

QY 242 KFSWLHASGVNVEDLSLDLTDVLENGSETCEANAVGITATGILTVAHPKFVI 301

Db 256 ----ITDESERFALRGSSLLISDVTEEDGVTYCIADENETIEQAELNAVQPFPELK 311

QY 302 RPKNQIIGVEDVLFPCQANGHPPTLYKSVDGSSLLPYGRDMGEVILTPEGRSVL 361

Db 312 RPANIATHESMDLFECEVTGCKPTPVYWKMGD-WVIP----- 349

QY 362 IARFAREDSGVYTCNALNAVGSYSSTRVSVSYDTPELPPPIEQSGVQWQLPWKSIIVL 421

Db 350 -----SOYFRVYKRNQLVLGLVKS----- 369

QY 422 PCLRTIGTPVPQSWYLGDIPDIQVEHERRNLSDAGALTISDLQHEDBSGLTICVASNRNG 481

Db 370 -----DEGFQCIAENDVG----- 383

QY 482 KSSMSG----YLRLLD-PTNPNIKFPRAFLSTYP--GP-GPKPMVEX-----GENSVLT 529

Db 384 NAQAQAQLILLDLDAIPLPPTSLSATNDHLAAPTGLPTAPRDVAIILTSRFPIRL 443

QY 530 SW-TSNEVHGSSL--VGVYTEMFGNETDGWAVGTNRQNT--PTQT--GLLPGV 579

QY 444 TWRTPTSPQODMUTSIFPTKYGINSIRE-----RVENTSRPGETQMIGNMPET 493

QY 580 NYFFLIARENHSHGLSPSMSEPIVTGTRYFNGDLSERASSLLSGDVELLSNASVSDS 639

Db 494 VYFPRVVAQMKHS--HESSAPLKVATQ----- 519

QY 640 TSKMLTWIIINKIVBGFVITYAROLPNPINNPAVTSNTNPLLSTSTSASASASAL 699

Db 520 -----PEVOLGPCA----- 528

QY 700 ISTKNIAAAGKRDETNOGGSGAPPPLMTKYRMILTINLGGG----- 741

Db 529 ---DNIRY-----AGSPSTVYVWE--TPLSGNGEIQNYKLYTMXGQDSDS 571

QY 742 ----ASCTITGLQYTLIEFFIVFPFYKVBEGKPSNRARIETLDVSEAPYGMALL 796

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

ON protein - protein search, using sw model

Run on: January 22, 2001, 12:27:18 ; Search time 162.41 Seconds
(without alignments)
274.602 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTFVKNDP.....RSLLSNSGSGTSSQPAHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
	1	597.5	8.2	1493	1	NEOL_MOUSE	P97798 mus musculus
	2	589.5	8.1	1377	1	NEOL_RAT	P97603 rattus norv
	3	575.5	7.9	1461	1	NEOL_HUMAN	Q92859 homo sapien
	4	568	7.8	1036	1	AXOL_CHICK	P28685 gallus gall
	5	560	7.7	1443	1	NEOL_CHICK	Q90610 gallus gall
	6	555.5	7.6	1284	1	NRCA_CHICK	P35331 gallus gall
	7	552.5	7.6	1447	1	DCC_MOUSE	P70211 mus musculus
	8	551.5	7.6	1447	1	DCC_HUMAN	P43146 homo sapien
	9	544.5	7.5	1040	1	AXOL_HUMAN	Q02246 homo sapien
	10	540.5	7.4	1040	1	AXOL_RAT	P22063 rattus norv
	11	538	7.4	2012	1	DSCA_HUMAN	Q06469 homo sapien
	12	536.5	7.4	1018	1	CONT_HUMAN	Q12860 homo sapien
	13	534	7.3	1239	1	NRG_DROME	P20241 drosophila
	14	531.5	7.3	1020	1	CONT_MOUSE	P12960 mus musculus
	15	521	7.2	2029	1	LAR_DROME	P16621 drosophila
	16	517.5	7.1	1010	1	CONT_CHICK	P14781 gallus gall
	17	511.5	7.0	1260	1	CAML_MOUSE	P1627 mus musculus
	18	509	7.0	1257	1	CAML_HUMAN	P32004 homo sapien
	19	497.5	6.8	1259	1	CAML_RAT	Q05595 rattus norv
	20	495.5	6.8	1897	1	PTPF_HUMAN	P10586 homo sapien
	21	487	6.7	1266	1	NGCA_CHICK	Q30696 gallus gall
	22	485.5	6.7	1912	1	PTPD_HUMAN	P23468 homo sapien
	23	457.5	6.3	1913	1	KMLX_HUMAN	Q15746 homo sapien
	24	441	6.1	3707	1	PGBM_MOUSE	Q05793 mus musculus
	25	427.5	5.9	4393	1	PGBM_HUMAN	P98160 homo sapien
	26	412.5	5.7	1070	1	PTK7_HUMAN	Q13308 homo sapien
	27	406	5.6	1091	1	NCAL_CHICK	P13590 gallus gall
	28	403.5	5.5	761	1	NCAL_HUMAN	P13592 homo sapien
	29	398.5	5.5	837	1	NCM2_HUMAN	Q15394 homo sapien
	30	396.5	5.5	853	1	NCAL_BOVIN	P18336 bos taurus
	31	389.5	5.4	837	1	NCM2_MOUSE	Q15336 mus musculus
	32	389.5	5.4	1115	1	NCAL_MOUSE	P13595 mus musculus
	33	389	5.3	1088	1	NCAL_XENLA	P16170 xenopus lae

34	388.5	5.3	725	1	NCAL_MOUSE	P13594 mus musculus
35	388.5	5.3	848	1	NCAL_HUMAN	P13591 homo sapien
36	388.5	5.3	1092	1	NCAL_XENLA	P36335 xenopus lae
37	385.5	5.3	858	1	NCAL_RAT	P13596 rattus norv
38	380	5.2	1906	1	KMLX_CHICK	P17799 gallus gall
39	368	5.1	1051	1	PTK7_CHICK	Q91048 gallus gall
40	356	4.9	1666	1	SKLM_MOUSE	Q62234 mus musculus
41	355.5	4.9	2481	1	UN52_CAELE	Q06561 caenorhabdi
42	346	4.8	811	1	FS22_DROME	P34083 drosophila
43	339	4.7	873	1	FS21_DROME	P34082 drosophila
44	323	4.4	1451	1	MYML_HUMAN	P52179 homo sapien
45	321	4.4	898	1	FAS2_SCHAM	P22648 schistocerc

ALIGNMENTS

RESULT 1
NEOL_MOUSE
ID NEOL_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEOL OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN.
RX MEDLINE=97407651; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RT expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UBQUITOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
CC OR SEND AN EMAIL TO license@sib-sib.ch).
CC -----
DR ENGL: Y09535; CAA70727.1; .
DR HSP: P02751; ITTG.
DR MGD: MGI:1097159; NEOL.
DR INTERPRO: IP0001777; .
DR INTERPRO: IP0003006; .
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; ig; 4.

CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53101; CAA44815.1; -
 DR PIR: S22128; S22128.
 DR PIR: S22383; S22383.
 DR HSP: P56276; ITLK.
 DR INTERPRO: IP0001777; -
 DR INTERPRO: IP0003006; -
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; lg; 6.
 DR Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1 23 OR 25 (POTENTIAL).
 FT CHAIN 24 1036 AXONIN-1.
 FT PROPEP 7 1036 REMOVED IN MATURE FORM.
 FT DOMAIN 49 113 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 143 211 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 249 308 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 336 397 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 428 490 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 518 589 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 599 608 HINGE (POTENTIAL).
 FT DOMAIN 601 607 GLY/PRO-RICH.
 FT DOMAIN 608 709 FIBRONECTIN TYPE-III.
 FT DOMAIN 710 811 FIBRONECTIN TYPE-III.
 FT DOMAIN 812 912 FIBRONECTIN TYPE-III.
 FT DOMAIN 913 1009 FIBRONECTIN TYPE-III.
 FT MOD_RES 724 724 BLOCKED.
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 493 493 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 520 520 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
 SEQUENCE 1036 AA; 113301 MW; 08800143B2779794 CRC64;

Query Match 7.8% Score 568; DB 1; Length 1036;
 Best Local Similarity 23.1%; Pred. No. 2.9e-23;
 Matches 250; Conservative 116; Mismatches 352; Indels 364; Gaps 40;

QY 4 PRIIEHMDITVPRK---NDPFTFCQAEQNPPTTIQWQDREGLATDGSHRIMLAGL 60
 DB 32 PVFEDPATLTPBGSAREKVLTCRANRPAITKWKNGTLMGQPS-RTYRLVAGDL 90
 QY 61 PFLXVLRHSRSDAGTWCEARNEFGVARSNATQVAVLRDFEPLR-----107
 DB 91 VISNPYKAK---DASSTQCVTARNRGTVYSRESLRGLQ-FBSAREDERPVKITEGTV 146
 QY 108 -----107
 DB 147 MFTCSPPPHPALSYNLLNEFPNFIADGRGFRVSTGTLNLIATKSDGLNCSFATS 206
 QY 108 -----PANTRVAGCEVALMGAGPGSPPEQ 133
 DB 4207 HIFDTIPSVFSRQSLAEDARQYASIRAKPADIYALQGMVTLCEA-PGMPVQ 265
 QY 134 ISNRK---NQUTLVNMGKRIIVDGNLAIQEARQSDGRTQCVVNVGTRESATFLK 191
 DB 266 IKNRKLDSQSTKWLSEPL-----LHIQVDFEDRTYCEAKING-RDTYQRII 317

QY 192 VHVRFPLINGPQQTAVGVSSVFPQCRIGDPLDPLVRRRTASGNNMLKRFWSHSAG 251
 DB 318 IHAQPMWLVDITDADIQSLDMSCVASKPRPAPVWLR---DQQLP-----ASN 366
 QY 252 RVVLEDRSLKLDVTDLMGEMGTCEADNAVGTATGILVTHA-PKPFVIRPKNQLVEI 310
 DB 367 RIEV-SGGELRFSKLVEDSGMQCVANRKGTVYASAEITVQALAPDFRLAPVRLIPA 425
 QY 311 ---GDVLPFCQANGHPRPTLYSVNVEGNSLLLPYRGRMEVTLTPGSRVLSIARFAE 368
 DB 426 ARSGVITPCPRAAPKATVLT--KGTELTNNSR-----VITADGTLILQ--NISK 476
 QY 369 DSGKVTCCNALNAGSVSSRTYSVDQFELPPEIIPGVQVTLVKSIVLPCRTGL 428
 DB 477 DEGR-TCYFANENYKANTSTGLSVDATK-----ITLAPSSADINVENLILQCHAS 530
 QY 429 PVQVY---SWLGDGIPIDVQHE---RRNLSDA-GALTSIDQ-RHEDGSLTYCVASNR 480
 DB 531 PTMDLFTTMSLDDFPIDLDKSGHTRASVKEAGVLAVMAQLK--SGRTCTAQTVY 588
 QY 481 GKSSWGLRGLDTPPTNPNKIFRPAELSTYGPGRPKPMVKENSVTLNWSNRKVGGS 540
 DB 589 DSTSESATLVKGP-----PGPGGVVVRDIDGTVQLSWSRGDF-WHS 631
 QY 541 SLUGVIV---MFGKNEIDGNVAGTVRQN-----TPTPTGTGLPGVNTYFLIRANS 591
 DB 632 PIARYSTIARILLNKK-----WQMTNPNVIEGNASTAQVNLIPMDYEVFVLSN 686
 QY 592 GLSLSPMSSEPTVGTGRVNSGLDSEARSLSGCVVLSNAGSVSDTSKMLTQWING 651
 DB 687 GVEPES-----692
 QY 652 KYVEGPTVYARQLPNIVNPAVPTSWNPLLGSTSTASASASALISTK---PNIA 708
 DB 693 -----LP-----SKRIETKAPTAP 709
 QY 709 AKGRDGETWQSGGAG-----TPLNTKYRMLTILNGG-----ASSCITGV 751
 DB 710 SGL-----GGGGAGNELIITNPTLADTQ-----NGDGPYILSPKKGITGSLTAR 756
 QY 752 -----QTLIEFIVFPVYSVKGKPSNRIATLEVDVSEAPTEGALL 795
 DB 759 PHAESLHVIVYRNESIGPTTFVFKATNRKGGSPSTALVYSAREZKVPAPFVATA 818
 QY 796 LNSSAVFLKAKPELKRHGLVNLVHY-IVRGDITAHNFSLRLNVDAASPTLVLAN 854
 DB 819 VLSSDMWSVSPVQDGTGVLGTEIYRWKDKDEKAAKRVRTAGLVTSARV-----GL 874
 QY 855 TBGVNVTGVGAAGNNGVPGIC-----VPATLRDLPITKRLDPPIN 896
 DB 875 NPNTKTHSVYRNAYRAGAGPPSPNTITTKPPRRPNTGSLWITLGSVTYIKMPVQA 934
 QY 897 RD 898
 DB 935 AD 936

RESULT 5
 ID NEOL_CHK
 AC 090610;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEOGENIN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LEBHORN; TISSUE-EMBRYONIC BRAIN;
 RX MEDLINE-95105243; PubMed-7806578;

us-09-540-245a-16.rsp

Page 6

```

Qy 128 GSPEPQISIKKQKQITVWKKRIRIVDGNALQIARQSGDCGVCQVWKNVGTRESAT 187
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 143 VDLAPFVFWGDDQRFSL--DWRVFLPGKALLGNATDGGFVTRVIESGGGPKYSE 200

Qy 188 AFLKVVHVP-----FLIRGQWQTVAGVSSVFGCRIGGDDPDVWARRRATSGMGLPR 241
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 201 AEIKLILPDPEEQSLVFWVRQSSLLVFGQNAVEFCVAGGFPPIVYV--FANGEE-- 255

Qy 242 KFSWLHSHASGRVWVLEDRSLKDDVLEDMGECTCEADNAVGGITATGILTVHAPPFVI 301
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 256 ---ITDESERFALRAGGSLNISDVTEEDVGVTICTADENETIEAQAELAQVQVFFLK 311

Qy 302 RPKKVLVEIGDEVLPFCQANGRPPTLVSVGCSNGLLPGTRGMGVTITLTPGSRVLS 361
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 312 RPNATIHESMDVFECEVTGKPIPTTKWVKGD--VVP- 349

Qy 362 IARFAEDSGAVVTCTMALNAVGVSSSRVTYSVDTFQLPPIIBQGVWQTLPKSVILV 421
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 350 ----SDTPKIVKHEHNLVLQVLSK- 369

Qy 422 PCRTLPTGVPGVSNWLDGIPIDVQHEHRRNLSDAGALITISDLQHEDEGLTVCSAVNRG 481
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 370 ----DESPITQIADNQG- 383

Qy 482 KSSNSG---YRLDIT--PTNPNIKFRAPELSTYP--GP-PGKPMVEX--GENSVIT 529
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 384 NAQAQAQLILLDLDAVPLPPTSLISATNDLADATTPGTLPTRADVATLVSTRFIRL 443

Qy 530 SW-TRSNKVGSSSL--GVGVEMFGNKEDGWAVGTRVQNT--PTQT--GLLPGV 579
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 444 TWRTVSPDQGNLTLSISPYTEKINRE-----RVNSTRSPGQVQINQIMPET 493

Qy 580 NYFFLIRANSHGSLSPMSPEPITVGTFRFNSCLDISEARSLSCDVELSNVSDWS 639
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 494 YVYFVPAVQNRG--HGESSAPLKVATQ- 519

Qy 640 TSMKLTVLIINGKIVBGFYVYARQLVNPVLPVNPAPVTSINTPLGSTSTASASASAL 699
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 520 -----PEVQLDGA- 528

Qy 700 ISTKPIAAAGKRGDETQSGGGAPPLNTKYRMLTILNNGG- 741
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 529 ---PNIRAT-----AGSPSTVYVW--TPLSGNGEIQNTKYLTYEKQGDSQF 571

Qy 742 ----ASSCTITGLVQTYLTFEYFFVPIPKSVBKPSNRARTLEDVPEAPYGEALL 796
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 572 DVNDVAGLSYTLTKLKITYESRFVATNKNHKGVGSTQDVVTRLSDFVSAAPONLEAR 631

Qy 797 NSSAVFLKKPAPELKDHRGVLLNVHVRIGDITAHNFRILTNWTDIASPTL-VLANL 854
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 632 NKSINLHQLWQPPACTHSGQITGIRKYRYSRK-----SDVTSVGQTLQLIEGL 684

Qy 855 TEGVMTVGVGAAGNAGVGPYC-----VPATRLDPI 886
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 ERGTENFRIMAMVINGTPAGDWMASATFESDLDSESRVPEVSSLVLRPL 735

```

```

RESULT      6
NRCA_CHKCK
ID  NRCA_CHKCK      STANDARD;      PRT; 1284 AA.
AC  P353131;
DT  01-FEB-1994 (Rel. 28; Created)
DT  01-FEB-1994 (Rel. 28; Last sequence update)
DT  15-JUL-1994 (Rel. 38; Last annotation update)
DE  NC-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RS  STRAIN=WHITE LEGHORN; TISSUE=EMBRYONIC BRAIN;
RX  MEDLINE#P=31258407; PubMed#2045418.

```

[illegible]

RA Grumet M., Mauro V., Burgoon M.P., Edelman G.M., Cunningham S.A.;
 RT "Structure of a new nervous system glycoprotein, Nr-CAM, and its
 RT relationship to subgroups of neural cell adhesion molecules.";
 RL J. Cell Biol. 113:1399-1412(1991).
 RN [2]
 RP SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-EMBRYONIC BRAIN, AND RETINA;
 RX MEDLINE=9238110; PubMed=1512236;
 RA Kayen J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;
 RT "Bravo/Nr-CAM is closely related to the cell adhesion molecules L1
 RT and Nr-CAM and has a similar heterodimer structure.";
 RL J. Cell Biol. 118:1259-1270(1992).
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
 CC NEURITES, ETC. SPECIFICALLY INVOLVED IN THE DEVELOPMENT OF OPTIC
 CC FIBRES IN THE RETINA.
 CC -1- SUBUNIT: HETERODIMER, COMPOSED OF AN ALPHA AND A BETA CHAIN.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: RETINA AND DEVELOPING BRAIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING NEURAL RETINA AND
 CC EMBRYONIC BRAIN TISSUE.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 9 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: X58482; CA641391.1; .
 DR EMBL: L08960; AAA8632.1; .
 DR HSP: P20241; ICFB.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 5.
 DR PFAM: PF00047; Ig; 6.
 DR PRINTS: PRO0014; PNTPEI11.
 KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
 KW Transmembrane; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 1284 NG-CAM RELATED CELL ADHESION MOLECULE.
 FT DOMAIN 25 1143 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1144 1166 POTENTIAL.
 FT DOMAIN 1167 1284 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 56 125 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 155 220 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 261 323 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 351 415 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 445 508 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 536 599 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 638 699 FIBRONECTIN TYPE-III.
 FT DOMAIN 738 799 FIBRONECTIN TYPE-III.
 FT DOMAIN 837 906 FIBRONECTIN TYPE-III.
 FT DOMAIN 943 1006 FIBRONECTIN TYPE-III.
 FT DOMAIN 1057 1114 FIBRONECTIN TYPE-III.
 FT DISULFID 63 118 POTENTIAL.
 FT DISULFID 162 213 POTENTIAL.
 FT DISULFID 268 316 POTENTIAL.
 FT DISULFID 358 408 POTENTIAL.
 FT DISULFID 452 501 POTENTIAL.
 FT DISULFID 543 592 POTENTIAL.
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 581 581 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 595 595 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 834 834 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 969 969 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 995 995 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1059 1059 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPIC 612 621 MISSING (IN ISOFORM AS10).
 FT VARSPIC 1027 1038 MISSING (IN ISOFORM AS12).
 FT VARSPIC 1039 1131 MISSING (IN ISOFORM AS93).
 FT VARSPIC 1202 1205 MISSING (IN ISOFORM AS-CYT2).
 FT CONFLICT 209 209 Y -> E (IN REF. 2).
 FT CONFLICT 680 680 H -> Q (IN REF. 2).
 SQ SEQUENCE 1284 AA; 141851 MW; A3570BF9C3D470P CRC64;

Query Match 7.6%; Score 555.5; DB 1; Length 1284;
 Best Local Similarity 20.8%; Pred. No. 1.8e-12;
 Matches 277; Conservative 179; Mismatches 515; Indels 35; Gaps 51;

Qy 8 EHPDTPVKNDFPFQCAQENPTPTQWKGDELKLTQDGERIMLPAGGLFLKVIH 67
 Db 46 GSFDDIVDPANIVQCAKAPPSPSWKNGTBFDDKQAVYKAPNGSLVWIMM 105
 Qy 68 S-RRESAQTWCARKEFGVARSNRLQV--AVLRDFEPLRANVQAQVALMEQC 124
 Db 106 GVAREDAVGQYCTARNRGAISNIVTPRSPLTKKLEPLNHRV--GGSLVLCNR 163
 Qy 125 APAGSPPEQISWRKNGQTLNWKGRIVRVDGNAIQARQSDGGR--TQCVW----- 176
 Db 164 PVPVGLPPPIIPMDNA-FQRLQPSQSVSGGLNGDLIFSIV-QPEDTVDYCTIARFNTQ 221
 Qy 177 ----KVVGTRESATPLKVRVRFPLIRG--NQTVAVGSSVVFQCRIGGDPLOVL 228
 Db 222 TIQKQPIQSVKFSYK--PYTERFVLLTPMGSTSKVRLGNVLLCEALGALPPTVR 279
 Qy 229 WRTASGGRNPLKRTSLWLSASGRVHLEDSRLKLDVLTDMGCTTCEADNVGGITAT 288
 Db 280 WIK--BGGLPANKTFPFN-----KXLIKIDVSEADSGNCTKANTLGSTHV 328
 Qy 289 GILTVHAPKPIYRPFKQLVEIDGLVFLQCAQNGHPRPLTVSVSGNSLLPLRGDRM 348
 Db 329 ISVTVKAPPTITAPRVLSPGEGDTLICRANGNKPISWLNGLVPIATP--EDPSR 386
 Qy 349 EVILTPBGRSLVLSARFARSDGKVTYCNALNAGSVSRVTVSDVQFELPPIIEQG 408
 Db 387 KV-----DGTITFSA--VQERSAVQCCNANSGYLLANFVNLAE--PRRLT--P 435
 Qy 409 VQTVLP--KSIIVLPCLRTGLPVPQVSWYGLDIPDVQEHERRNLSDAGALTISDLRH 466
 Db 436 ANKLIVQIASPALDCAITGFSKPKPEIDF--RGVGSILRGNTEVFEDNGTLEIPVAK- 493
 Qy 467 EDEGLTCTVSNRNGKSSNGVLRDLTPT-----NPMKFFFA-----PEL-- 507
 Db 494 DSTGTTCTVANKLGTQNEVQLVKDPTMIKQPKYQISQAASFQCVKHPDPLIP 553
 Qy 508 -----STY----- 510
 Db 554 TVIWLKDNNEPLPDERFLVGDNLIMVTVKDDGTYTIVNTLDSVSASALIVVAAP 613
 Qy 511 -----PDPKPKQWKEGNSVTLKSRKNGKSSVGVGVIM--PKNEDGQWA 560
 Db 614 PPTAIVARPPPLDLELTQQLSRISZLWPGCE--NNSPTNFVIEYDGLHGPQW-H 671
 Qy 561 VGTGVN--TETQGLLPGVNTFFLIRANSNGHLSLSPMSPESTVGTIRFNSGLDSEA 619
 Db 672 VQTEVPGSHTVQLKLSPPYNTSFVIANRIGRSPSPSEF-----QTLTKSANPDE- 724
 Qy 620 RASLLSGDQVWLSASVSDS--TSMKLWQINGIKYEG---FTYVTRALPNIWNPA 673

FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.

-1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION AND REMAIN AT THIS LEVEL IN THE ADULT.

-1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

-1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X85788; CAAS59786.1; -.
DR HSP: P56276; 1TLK.
DR MGD: MGI:94869; DCC.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR009006; -.
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; Ig; 4.
DR PRINTS: PR00014; FNTYPEIII.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, LONG ISOMFORM.
FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, SHORT ISOMFORM.
FT INIT_MET 85 85 FOR SHORT ISOMFORM.
FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1098 1122 POTENTIAL.
FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 54 124 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 154 219 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 317 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 819 838 MISSING (IN EMBRYONIC MM).
SQ SEQUENCE 1447 AA; 158299 MW; 0D1F1097C22D59FF RC654;

Query Match 7.6%; Score 552.5; DB1; Length 1447;
Best Local Similarity 22.04%; Pred. No. 31e-21
Matches 316; Conservative 172; Mismatches 532; Indels 415; Gaps 66;

Gy 7 IEHPDPTVANDPFTPNQAGNG-PPTTIGKDKREGLTKTGSRHMLGAPGGFLFKV 65
| | | | | | | | | | | | | | | | | | | | |
dbb 43 VSEPDAVTNPGNVLLNCASRDGRGVPIKKWKLALIALGMDRQQPLWSGLLIQNI 102

Gy 68 IHSF-RSDATYITCKAE-NFGVARSNNATLQA-VLRDRFPRLPEANTRVGQVALME 122
| | | | | | | | | | | | | | | | | | | | |

[illegible]

Qy	957	PSLSARAG-----GTWDSSTGGM--WRPFS--PGGSLSDMKQHIADYAPVACGA	1004
		: : : : : : : : : : : : :	
Db	1139	-----SKRGSKQKQLRPDLWIMHEEMXNIXETGCTDAGRSGSIQS--CQDLIPWSSG	1193
Qy	1005	GSPAGGPTTSGSGSGS-----GAGSGAGSGDDHGGHSGERNQRTVGEYSNIPDTAEVSSF	1060
		: : : : : : : : : : : :	
Db	1194	SETQMGSASGSGQDTEGAGSGSMTLERSLAARATRAKIMIPMAQGS--SNPVAFAT	1251
Qy	1061	GKAPSEYGRGHNGASAPATYTSLSLSPHQQQQQQQPRYQRPVP-----GTGLQRPMI	1112
		: : : : : : : : : : :	
Db	1252	VPVPLSAPYQGLPILPS--TCQTH-----PQTLRPVFPVPTLSVDVGAGAG--	1297
Qy	1113	PHYQQQHQHQQQQAQQTHQHQHQLHQHQLPSPNIXQMSTTSEYPTNTGSGRSV	1167
		: : : : : : : : : : : : : : : : : :	
Db	1298	-----TQSYSEGGPTQQOPLMPLA-----OPEPHFSSEASRPTITPACH	1336

RESULT 8

ID DOC_HUMAN
 AC P43146;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
 GN DCC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95011532; PubMed=7926722;
 RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
 RA Vogelstein B.;
 RT "The DCC gene product in cellular differentiation and colorectal
 tumorigenesis.";
 RL Genes Dev. 8:1174-1183(1994).
 RN [2]
 RP SEQUENCE OF 1-750 FROM N.A.
 RX MEDLINE=90100559; PubMed=2294591;
 RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
 RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
 RA Vogelstein B.;
 RT "Identification of a chromosome 18q gene that is altered in
 colorectal cancers.";
 RL Science 247:49-56(1990).
 RN [3]
 RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
 RX MEDLINE=91121517; PubMed=1991332;
 RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
 RA Oliner J.D., Kinzler K.W., Vogelstein B.;
 RT "Scrambled exons.";
 RL Cell 64:607-613(1991).
 RN [4]
 RP GENE STRUCTURE, AND VARIANT CARCINOMA HIS-1375.
 RX MEDLINE=94245241; PubMed=8188295;
 RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
 RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
 RT "The DCC gene: structural analysis and mutations in colorectal
 carcinomas.";
 RL Genomics 19:525-531(1994).
 RN [5]
 RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
 RX MEDLINE=94243823; PubMed=8187090;
 RA Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
 RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
 human esophageal squamous cell carcinomas and their relation to
 metastasis.";
 RL Cancer Res. 54:3007-3010(1994).
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -1- SUBCELLULAR-LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL

CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
 CC -1- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
 CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
 CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
 CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
 CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
 CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
 CC OR SEND AN EMAIL TO license@isb-sib.ch).
 CC -----
 DR EMBL: X76132; CAA53735.1
 DR EMBL: M32292; AAA35751.1
 DR EMBL: M32286; AAA52174.1
 DR EMBL: M32288; AAA52175.1; ALT_SEQ.
 DR EMBL: M32290; AAA52176.1
 DR EMBL: M53696; AAA52177.1
 DR EMBL: M53700; AAA52178.1
 DR EMBL: M53702; AAA52179.1
 DR EMBL: M53718; AAA52180.1
 DR EMBL: M53698; AAA52181.1
 DR PIR: A54100; A54100.
 DR PIR: A40098; A40098.
 DR PIR: A38442; A38442.
 DR HSSP: P56276; ITLK.
 DR MIM: 120470; .
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFM: PFM0041; fn3; 6.
 DR PFM: PFM0047; lg; 4.
 DR PRINTS: PR00014; FNTYPELII.
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KW Anti-oncogene; Disease mutation; Polymorphism.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC.
 FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1098 1122 POTENTIAL.
 FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 54 124 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 154 219 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 254 317 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 426 522 FIBRONECTIN TYPE-III.
 FT DOMAIN 525 618 FIBRONECTIN TYPE-III.
 FT DOMAIN 619 716 FIBRONECTIN TYPE-III.
 FT DOMAIN 722 816 FIBRONECTIN TYPE-III.
 FT DOMAIN 840 940 FIBRONECTIN TYPE-III.
 FT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
 FT DISULFID 61 117 BY SIMILARITY.
 FT DISULFID 161 212 BY SIMILARITY.
 FT DISULFID 261 310 BY SIMILARITY.
 FT DISULFID 352 400 BY SIMILARITY.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 168 168 M -> T (IN OESOPHAGEAL CARCINOMA).
 FT /FTID-VAR_003909.
 FT VARIANT 201 201 R -> G.
 FT /FTID-VAR_003910.
 FT VARIANT 1375 1375 P -> H (IN A COLORECTAL CARCINOMA).
 FT /FTID-VAR_003911.
 FT CONFLICT 138 138 MISSING (IN REF. 3).
 FT CONFLICT 233 329 MISSING (IN REF. 3).
 FT CONFLICT 421 421 MISSING (IN REF. 3).

SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766D0471F CRC64;

Query Match 7.6%; Score 551.5; DB 1; Length 1447;
 Best Local Similarity 21.8%; Pred. No. 3.5e-22;
 Matches 312; Conservative 177; Mismatches 542; Indels 397; Gaps 64;

QY 5 RIIEHPMDTVPKMDPTFFNCQARGN-PTPTIQWFKDGRLEIKTDGSRHMLPAGLIFFL 63
 DB 41 RFLSEPSDAVIMBQGNWLLDCSAESRQGVPIKMKKGDHLALDMDKQQLSNCSSLIQ 100
 QY 54 KYIHSR-RESIDAGTYCEAK-NEFGVARSNATLQVA-VLRDEFLPANTVAGQVEAL 120
 DB 101 WILSRHRRKPDGLYLQCEASLSDGSSIIISRTARVAVACPLR-PLSQTSVTAFMGDTVL 158
 QY 121 MBOCAPPGSPQLISWRKNGQTLN-LVGNKRIRIVDQGNLAQBARQSDGGRVYVW 179
 DB 159 LKCEV-IGEPVPIHWNKQDQTLPIPDGRVVLPSGALQISRLQPGDGIYKCSANP 217
 QY 180 VYRESATAPLX-----HYRFLRIGPQNTAVVGSVVFQICRIGDPLDPLMRHTA 233
 DB 218 ASSRTGNEAEVRLSDPLGRQLYLPQSPVVAIBGKDAVLCGVSTPPPTFWLAGE 277
 QY 234 SGNMPLKMSWLSASASRNVHLEBSLKDVLDTLDMGEYTCADNAVIGATIGTLTV 293
 DB 278 EVILQSKSKY-----LLGSSNLLISWTDSDSMTCVTKNENISASAEITV 327
 QY 294 HAPPKVPIRKNLQWEIGDEVLFECQANGPPPTLWSVGNSSLLPGTDRGMEVTLT 353
 DB 328 LVWPMPLHPSNLSATYESMDIEFCTVSGKVPVTHMKNGD--VLP--SDTFQIV-- 380
 QY 354 PBRSLVSLAFAREDSKVVTCNALNAVGSVSRVVSVDQFELPPPIEDGFWNQLT 413
 DB 381 -GGMLRLGVKSDGG-FYQCAVNNAGNAQT-----SAQLVFKPAIPSSVLESA 431
 QY 414 PVKSIIVPLKTLGTLPVQVQSVYLDGIPIDV-----QSEHR-----NLSAGA 457
 DB 432 PRDQVPL-----VSRFVLSW-----RPAEAGNIQTPTVFSRBDGNERANLTQGS 484
 QY 458 ---LISDLQRHDEBLTY-CVASNRNKSNSGSLRLDTPTNPIKFRAPFSLTTPG 513
 DB 485 LQLTVGNL-----KPEAMTFRVYVNEWGPG-----SSQIKVATQELQV-PGP 530
 QY 514 PGKPMQVXKSGSVLTSWTRSKVGVSSSLVGYIEM-----GRNETDGVWAGVTRQN 567
 DB 531 VEMLQAVSTPSILITWBPAYANG-PVQGY-RLCEFTVSTGKED-----NIEVDG 580
 QY 568 TIFTQTGLLPVNTYVILRAENSHGLSLPMSRPEITVTRTFNSGLDSEARASLSDG 627
 DB 581 LSLKLDGLKGLTSLRLFLAYRKY---PGVSTDDITVVT-----LSDVPSAPQV 629
 QY 628 VVELSASVSDTSNKLTV-----QIINKYVBGTV----- 659
 DB 630 SLE-----VWSRSIKYMWLPSPSQSGQ-FITGYIKRHKTRRGEMETLPPNWLTF 683
 QY 660 -----YARQLPNFVNNPAP-----VTSNT--NPL-----LGSSTSSASASALIS 701
 DB 684 TGLEKSGQSTFQVSNATVNGTGPNNVTAETPENDLSDQVDPQSSLHVBPQNCIN 743
 QY 702 T-----KPNIAAAGKQ-----ET-----NQSG 720
 DB 744 SWTPPLPNFVNYGVIIGVGSPYARTVRVDSKQRTYSIERLESSHSVISLAFNAG 803
 QY 721 GGAP-----TPLE----- 727
 DB 804 BEVPLYESATSRITSDTPDPTVDYPLDDPTSPVPLSPMPGVQVAVLTHDRAVRS 863
 QY 728 -----NFKYRMLTI-----LNGGAS-----SCITGLVQITLIEVTFVFP 763
 DB 854 WADNSVQPKMKTSEVRLTYVWRMTSFSASAKYSEDTSLSLTATGAPKNTWTFESVMT 923
 QY 764 YKSVEGKPSNRIARTLEDVPSAPYTMGALLNS--AVFLKAKPELDRHGVLYNH 821
 | : : | | | : : : | : : | : : |

RESULT 9
AXOL_HUMAN
ID AXOL_HUMAN STANDARD; PRT; 1040 AA.
AC Q02246;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL
DE GLYCOPROTEIN 1).
GN TAXI OR TAG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE-BRAIN;
RA MEDLINE=93145965; PubMed=8425542;
RA Hasler T.H., Rader C., Stoeckli E.T., Zuellig R.A., Sonderegger P.;
RT "cDNA cloning, structural features, and eucaryotic expression of
RT human TAG-1/axonin-1";
RL Bur. J. Biochem. 211:329-339(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=94140354; PubMed=8307567;
RX Tsiotra C.P., Karageorgos D., Theodorakis K., Michaelidis M.T.,
RA Modi W.S., Furley J.A., Jessel M.T., Papanthakis J.;
RT "Isolation of the cDNA and chromosomal localization of the gene
RT (TAG1) encoding the human axonal glycoprotein TAG-1";
RL Genomics 18:562-567(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collab
CC between the Swiss Institute of Bioinformatics and the EMBL outst
CC the European Bioinformatics Institute. There are no restrictions
CC use by non-profit institutions as long as its content is in
CC modified and this statement is not removed. Usage by and for com

[illegible]

```

RESULT 10
AXOL_RAT
ID AXOL_RAT STANDARD; PRT; 1040 AA.
AC P22053;
DT 01-AUG-1991 (Rel. 19; Created)
DT 01-AUG-1991 (Rel. 19; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1).
GN TAXI.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC TISSUE=SPINAL CORD;
RX MEDLINE=90199890; PubMed=2317872;
RA Furley A.J., Morton S.B., Manalo D., Karagozev D., Dodd J.,
RA Jessell T.M.;
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily

```

member with neurite outgrowth-promoting activity.";
 Cell 61:157-170(1990).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
 CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
 CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
 CC -1- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS, AND IN ADULT
 CC BRAIN, SPINAL CORD AND CEREBELLUM.
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED ON A SUBSET OF AXONS
 CC IN THE DEVELOPING RAT NERVOUS SYSTEM.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN-TYPE III-LIKE DOMAINS.
 CC
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
 CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS OF
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
 CC OR SEND AN EMAIL TO license@isb-sib.ch).
 CC
 DR EMBL: M31725; AAA42201.1; .
 DR PIR: A34695; A34695.
 DR INTERPRO: IPRO01777; .
 DR INTERPRO: IPRO03006; .
 DR PFAM: PF00041; fn3; 4.
 DR PFAM: PF00047; ig; 6.
 KW IMMUNOGLOBULIN DOMAIN; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1 30
 FT CHAIN 31 71015 AXONIN-1.
 FT PROPEP 71016 1040 REMOVED IN MATURE FORM.
 FT DOMAIN 56 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 218 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 315 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 343 404 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 435 497 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 525 596 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 608 614 GLY/PRO-RICH.
 FT DOMAIN 613 708 FIBRONECTIN TYPE-III.
 FT DOMAIN 716 811 FIBRONECTIN TYPE-III.
 FT DOMAIN 818 910 FIBRONECTIN TYPE-III.
 FT DOMAIN 911 1005 FIBRONECTIN TYPE-III.
 FT SITE 796 798 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 206 206 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 479 479 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 527 527 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 777 777 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 832 832 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1040 AA; 113042 MW; 6870786514C24FB CRC64;

Query Match 7.4%; Score 540.5; DB 1; Length 1040;
Best Local Similarity 22.8%; Pred. No. 8.7e-22;
Matches 261; Conservative 160; Mismatches 442; Indels 283; Gaps 51;

[illegible]

RESULT 11
DSCA_HUMAN
ID DSCA_HUMAN STANDARD: PRT: 2012 AA

060469; 060468;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE PRECURSOR (CHD2).
GN OSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=98087574; PubMed=9426258;
RY Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
LYONS G.E., Korenberg J.R.;
RT "DESCAM: a novel member of the immunoglobulin superfamily maps in a
RT down syndrome region and is involved in the development of the
RT nervous system.";
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RY Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
Rosenthal A., Kodoh J., Shibuya K., Kawasaki K., Asakawa S.,
Rintaant A., Sasaki T., Hagihara K., Mitsuyama S., Antonarakis S.E.,
Shimoshina S., Shinzui N., Worsdick G., Bornischer K., Brandt P.,
Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
Ramses J., Beck A., Klages S., Hennig S., Riesselmann L., Deyand E.,
Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
Lehrach H., Reinhardt R., Raspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP FUNCTION.
RY Agarwala K.L., Wakamura S., Tsutsumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule OSCAM mediates intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
CC -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF023450; AAC17957.1. -.
DR EMBL: AF023449; AAC17956.1. -.
DR EMBL: AL163283; CAB90454.1. -.
DR EMBL: AL163282; CAB90436.1. -.
DR EMBL: AL163281; CAB90444.1. -.
DR MIM: 602523; -.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF000041; fn3; 6.
DR PFAM: PF000047; Ig; 9.
DR PRINTS: PR00004; FNTYPEIII.
KW immunoglobulin domain; glycoprotein; Signal; Cell adhesion; Repeat;

QY 269 EDGHEGTEADNAVGGITATG-ILTVHAPKPVIRPKNLVGEIDVLFQDQANGPPT 327
 Db 310 DDAAGTTCIDVSGVNGAQSFSLIIVNSVPTFKPEIATAEDZEVVFCRAQGPPEK 369
 QY 328 LYNVSEGNSSLLPGLYDRGMVLTPEGR---SVLSIARFADSDSGKVYTNALMAV 382
 Db 370 ISW-----IHNGKPIQSTPNPRTVDMIRIILVKGDTGN-TGCAATNSL 416
 QY 383 GVSSTRTVYSVDQFELPPPIEGQPVNQLPVKSIVLPCRTITPVPQVSW---YL 437
 Db 417 GYVYKQVILVQAE---PPTISEAPAAVSTVGRNVTIKCRVSGPKPLVKNLRAKML 472
 QY 438 DGIPIDVQHEHRRNLSDAGALISDLQHEDEGLYTCVANSRNGKSSGFLRLDTP? 495
 Db 473 TG-----GRYVQANGDLEIQDV-TFSDAGKTYCYAQNKFQIQADGLSVKEHRI 523
 QY 496 -WPNKIFPAPELSTY-----PGP-----PGKQWVEKGENSVLWS 531
 Db 524 TQEPQVYVAAGQSATFRCEAHDDTLEIDWKKQSDIDEAQPFPVKTNDNSLTAK 583
 QY 532 T-----RSN-----KVGQ-----SS 541
 Db 584 TMELDSEYTCVARTLDEATARANLIVQDVNPAPKLTGITQADAKRIEWDQGNRSP 643
 QY 542 LVGVVIEMPGKRLDGVWAVGRVQVITFT-QGLLPGVNTFFLIARNSHGLSLSPMS 600
 Db 644 ILRTITQFNTSTPASNDAYEKVPNTDSFVQVMSFWNTTFRVIAVWIGASPPSARS 703
 QY 601 EPIVTVGTRFVNSGLDSEARSLSDGVVLSNASVDSSTSMKLTQIINGKIVGFPYV 660
 Db 704 DSCSTQDP-----DVFFKNPDNVQGTPEPNVLSWSPFELHNPAPNFT---YVSW 753
 QY 661 ARQLPWPVYNNPAPVNTNPLGLSTSTASASASALSTXPNIAAAGKDEGTEN--- 717
 Db 754 KRDI-----PAAWENNIN-IPDWQNNIVADQPTFKVLKLVVAILDR-GENVAA 803
 QY 718 -----QSGGAP--TPLNTKYRMLTLINGGAGSCCTIGLVQVLYEFFVITPVSEVG- 769
 Db 804 EEVVGTSGEDRPLDPAFTMQRTISSTSGYMWNTVSEESVRGHGKTIQVWENGE 863
 QY 770 -----KPSNSRIARILE-----DVPSEAP---TGM 791
 Db 864 EGLREHVKGDTNHALVTQFKPDSANTARILANGFNGPVSAPPIDFTPEGVFSVQGL 923
 QY 792 EALLNNSAVFLKWKAPLKDQHGVLVHYVIGDITAHNFSLINLWIT-DAASPLV 850
 Db 924 DAIPLGSAFHLWKAKPLTP--NKLITGTYKTYEEVKVSVGERREIDPHITDPAVTKM 981
 QY 851 LANLITGMYTVGVAGNANAGVP---YCVPAIRL 882
 Db 982 MAGLKPNSKTRISITATKMGESSEHYEIKTIL 1015

RESULT 14

CONT_MOUSE

ID CONT_MOUSE STANDARD; PRT: 1020 AA.

AC P12960;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CONTACTIN PRECURSOR (NEURAL CELL SURFACE PROTEIN F3).

GN CNTNL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=BRAIN;

RX MEDLINE=89340657; PubMed=2474555;

RA Gennarini G., Cibelli G., Rougon G., Mattei M.-G., Goridis C.;

RT *The mouse neuronal cell surface protein F3: a phosphatidylinositol-

RT anchored member of the immunoglobulin superfamily related to chicken

RT contactin.*;

RL J. Cell Biol. 109:775-788(1989).

CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
 CC SYSTEM DEVELOPMENT.

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.

CC -1- MISCELLANEOUS: F3 SHARES WITH L1, N-CAM, MAG, AND OTHER CELL

CC ADHESION MOLECULES FROM NERVOUS TISSUE THE L2/HNK-1 CARBOHYDRATE
 CC EPITOPE.

CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: X1943; CA33075.1; -.

DR PIR: S0594; S05944.

DR HSP: P40189; 180Q.

DR MGD: MGI:105980; CNTN1.

DR INTERPRO: IPRO01777; -.

DR INTERPRO: IPRO03006; -.

DR PFAM: PF00041; fn3; 4.

DR PFAM: PF00047; Ig; 6.

KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;

KW Cell adhesion; Repeat.

FT SIGNAL 1 20

FT CHAIN 21 ?

FT PROPEP 7 1020

FT DOMAIN 58 121

FT DOMAIN 151 218

FT DOMAIN 256 317

FT DOMAIN 345 398

FT DOMAIN 429 491

FT DOMAIN 519 592

FT DOMAIN 604 611

FT DOMAIN 611 712

FT DOMAIN 713 814

FT DOMAIN 815 910

FT DOMAIN 911 1006

FT DISULFID 65 114

FT DISULFID 158 211

FT DISULFID 263 310

FT DISULFID 352 391

FT DISULFID 436 484

FT DISULFID 526 585

FT CARBOHYD 208 208

FT CARBOHYD 258 258

FT CARBOHYD 338 338

FT CARBOHYD 457 457

FT CARBOHYD 473 473

FT CARBOHYD 494 494

FT CARBOHYD 521 521

FT CARBOHYD 593 593

FT CARBOHYD 935 935

SQ SEQUENCE 1020/AA; 113388 MW; 9DCDA40BAACBCT CRC64;

Query Match 7.3%; Score 531.5; DB 1; Length 1020;

Best Local Similarity 22.4%; Pred. No. 2.6e-21;

Matches 234; Conservative 152; Mismatches 361; Indels 299; Gaps 45;

QY 1 GENPRIIEHPDITVPKND---PFTFNQAGNSPTPIQWFGDKRLKTDGSHRLMPLA 57

Db 38 GFGPIFEEQDINTIYFEESLEGVSLNCRASPSPPVYKNNRNDGV--DLNDRYSWVG 95

QY 58 GGLFVLKIVSRSDAGTYWCAKNPFGVARSNNATLQVAVLEDFLEP-ANTRWAGS 116

Db 96 GNLV---LNNPKQKQAGVYCIASNNQVGRSTELTSGFLT-DPPFFPEEPVVKVKG 151

QY 117 EVALMECCAPRSGSP-QISNWRNQTLLVGNKRIKRV--GGNLAIQEARQSDGQRY 173

Db 152 KGMVLLCDPTFFPDLLSYRWLNNEFPVITMDKR-RFVSQTNGNLIAVSSDSRGNTS 210
 Qy 174 CVVKNVGTRESATFLKVVHVPFLIRNGQNT-----AVGVSSVFFCR 218
 Db 211 CFVSSPSITKSVSKFIP-----LIPPERTTKPADIVVQPKDITMONGVNWLECF 264
 Qy 219 IGGDPLDVLNRTASGNGMPLKFSWLSASGRVHVEDRLSDLOVDLEMGRTCEA 278
 Db 265 ALGNPFTVIRKRVLE-----PMPSTAEI-STSGAV-----LXIFNQLQEDGLTECEA 312
 Qy 279 DNAVGIGATGILITHAPKFKVIRPKNLVBEIGDVEFLCQANGRPPLTMSVSGNSL 338
 Db 313 ENIRGKQKQARITVQAFPEWVEHINDVEIGSDLYMPCITATGKPIPTIRWLNKGTSS-- 370
 Qy 339 LLPGTRDGRM-----EVTILPEGRSVLSIARFARSDSKVYTCNALNAGVSVSRT---VVS 392
 Db 371 ----YRKGLRLYDVT-----ENAG-MYQCIABNAYGSIYANBAKLIA 410
 Qy 393 VDTQFPPPIIRQGGPVNQTLPVK-SIVVLPCRLGTPPVQVW-----YLDGIPDIQWE 446
 Db 411 LAPTFEMN-----MKKILKAANGVRIIECKPKAAPKPKSVSKQTEWL-----VN 457
 Qy 447 HERHNLSDAGALTISDLQREDEGLTVCVANSRNGKSSNGSLYDLPTPNWKEFPRAE 506
 Db 458 SSRILLWEDSGLEINNTIRN-DGCIYTCFANRNGRANSTGLVITNPT-----RILLAPI 512
 Qy 507 LSTPGPPGPKQWKEGENS-----VLSWRSKNVGGSSLVGIVIMRGN 553
 Db 513 NAD-----ITVGENNMQCAASFDPAIDLTFW-----SFGYVID-FMKE 552
 Qy 554 -----ETDGNVAVGTRVQNTITPTGGLPVNYPFLIRAENSRLGSLPMSSE 601
 Db 553 ITHITQRFNMWDANGELL-----INNAQLKAGRTICTAQTIVONSSASADLVIRGPPG 608
 Qy 602 PITVGTFRPNSGLDSEARSLSGCVVLSNVAVSVDTSKMLTKLQINGKIVYEGFVYA 661
 Db 609 P-----GLRIEDIRA-----TSVALTWS--RGS----- 630
 Qy 662 RQLPNFVPIVNPAPVTSNT-----NPLGSTSTASA----- 692
 Db 631 -----DNHSPISKYTIQTKTILSDWKDARTDPIIEGMSAKANDLVIRMEYFR 682
 Qy 693 SASASALISTKPNIAAG-KRDEGT-----WQSGGAP-----PLNTRYK---- 732
 Db 683 VVATNITLGTGEPSPISNRKTDGAARNVAPSDVGGGGTNRKELITWAPLSREYHGNF 742
 Qy 733 -----MLTILNGAGGASCTTIG-----LVQTTLYEYFIVPTKSVBKGPSNSRI 776
 Db 743 GIYVAFKPDGEKWKVYVTFNPTGRYVHKDETMTPTSAQVVKVAFNNKGOGPYSLAV 802
 Qy 777 ARTLEDVPSAPTGEMALLNSAVFLKWKAPKLDKRGVLLNHYVIRGIDT----- 829
 Db 803 INSAQDAPSEAPTEGVKVLSSSEISVHK-----VLEKIVESTQIRIWA 848
 Qy 830 AHNFHSFILTNTVIDASPTILVLNLTGVMGYTVAGVAGNNAVGPICPATILRLDPTIKR 889
 Db 849 GHDKCAAAHVRQVTSQEYARLENLLPQTYFIVGACNSAG---CGPSSVDIETPTX 904
 Qy 890 LDP-----FINQRDHY 900
 Db 905 APPSQPPIIISVRSRGTIITWDHV 930

RESULT 15
 LAR_DROME

ID LAR_DROME STANDARD; PRT: 2029 AA.
 AC P16621;

DT 01-AGO-1990 (Rel. 15, Created)

DT 01-AGO-1990 (Rel. 15, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE PROTEIN-TYROSINE PHOSPHATASE DLAR PRECURSOR (EC 3.1.3.48) (PROTEIN-TYROSINE-PHOSPHATE PHOSPHORYLASE).

GN LAR.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 LI [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90045860; PubMed=2554325;
 RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
 RT "A family of receptor-linked protein tyrosine phosphatases in humans and Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON-S;
 RX MEDLINE=96178473; PubMed=6598047;
 RA Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S., Saito H.;
 RT "The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in Drosophila.";
 RL Cell 84:611-622(1996).
 CC -1- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
 CC IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY (PI3PASE). IT CONTROLS MOTOR AXON GUIDANCE.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND PIONEER NEURONS IN THE EMBRYO.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M27700; AAA28668.1; -.
 DR EMBL: U36857; AAC47002.1; -.
 DR EMBL: U36849; AAC47002.1; JOINED.
 DR EMBL: U36850; AAC47002.1; JOINED.
 DR EMBL: U36851; AAC47002.1; JOINED.
 DR EMBL: U36852; AAC47002.1; JOINED.
 DR EMBL: U36853; AAC47002.1; JOINED.
 DR EMBL: U36854; AAC47002.1; JOINED.
 DR EMBL: U36855; AAC47002.1; JOINED.
 DR EMBL: U36856; AAC47002.1; JOINED.
 DR PIR: A36182; TDFFLK.
 DR HSP: P28827; 18PM.
 DR FLYBASE: Fbgn0000464; Lar.
 DR INTERPRO: IPR000242; -.
 DR INTERPRO: IPR000387; -.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR002006; -.
 DR PFM: PF00102; Y_phosphatase; 2.
 DR PFM: PF00041; fn3; 9.
 DR PFM: PF00047; Ig; 3.
 DR PRINTS: PRO0014; FNTYPELII.
 DR PRINTS: PRO0700; PRTYPHPTASE.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
 KW Cell adhesion; Immunoglobulin domain; Duplication.
 FT SIGNAL 1 32
 FT CHAIN 33 2029 PROTEIN-TYROSINE PHOSPHATASE DLAR.
 FT DOMAIN 33 1377 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1378 1402 POTENTIAL.
 FT DOMAIN 1403 2029 CYTOPLASMIC (POTENTIAL).

QY	4	PIIIEHIEFVTKNDPFTFVQAEQNPPTIQWQDGRLEATDTGSRIMLPAGGLF	63
Db	36	PEIIIRKQGVQVGVGASTYCAARGPPPSVWKKWKKVSGTQSRITVLQPGGIS	95
QY	64	KVTHSRRESADWTGYCEAKNEFQVARSNATLQVAVLEDFLEPA-----NTRVA	114
Db	96	RIEPVRAGDADAFCEAVENGVGVSADATLITY-----EGDTPAGFVITQGGFTRI	151
QY	115	Q-GYVALMDCGAPGSGPEPQISWRKNQTLNWKNRIRVDGNLAIQEARQSDGRTQ	173
Db	152	EWGHTVLTACAGTAPNITLWIKN-QTKVMNSPRLSKD-GFQIENSRDEQKTE	208
QY	174	CVVKNVGVTRESATAPLKVHRV-----PFLIRGPQVATV-VGSSVTFQCRIGDPLDLWM	229
Db	209	CVAKNSKTHSKATNLVYKVRVPPFTFSRPPTESISVMLGNSMLNSCIAVGSMPHVK	268
QY	230	RTTASG-----GNMLPKFWLHSAGRVHLEDRSLKDTDTLEMGCTTCADNAVGI	285
Db	269	MKGSDELPEKNZPI-----GR-NVLQ-----LINTQESANTICAASTLGOI	310
QY	286	TAGTLITVHAPPKAVIRKPN-QLVEI-GDEVLEQCAQGNRPPLTWISVE-----	333
Db	311	DSVSYVQVQSLP-----TAPDITQISEVTSATSLWENSYG-PEDIQITVYIQKPNANQAF	366
QY	334	GNSSLPGPDGRCMEVTLTPEGRSVLSIARFARSDSGKYTCNA-----LNAVGSVSR	388
Db	367	SEISGIITMTVTRALSPITTEYEFVTIWNANIGRPPSAPATCTGETKMSAPNPNQVR	426
QY	389	TVTSVDTQFELPEPIIEQGVQVQZLPKVSILVPCRLTGTPVQVSWNLGDISVDQREE	448
Db	427	TLSSSTWITWEPN-----ETPMQVGTQKVYY-----TNSNQPEASN-----N	456
QY	449	RRLSDAGALITSLDQREHDEGLTYCVASNRNKGSSNGSLRLDTPTPNFIKFRAP-EL	507
Db	467	QOWDNSELSTLTVSDVTH-----AIYT-----VYVQATISNGAGPSTMVPOV	508

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:51:13 ; Search time 559.88 Seconds
(without alignments)
289.105 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTPKNDP.....RSLLSNGSGCTSSQPAHNV 1381

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Aligned: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPREMBL_15.*
1: sp_archoea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	No.	Score	Match	Length	DB	ID	Description
1	3895.5	53.6	823	5	Q9VQ10			Q9vq10 drosophila
2	2462	33.9	859	5	Q9VP26			Q9vp26 drosophila
3	1790.5	24.6	1395	5	Q9W213			Q9w213 drosophila
4	1786.5	24.6	1395	5	044924			044924 drosophila
5	1498	20.6	1651	4	Q9Y6N7			Q9y6n7 homo sapien
6	1489.5	20.5	1612	11	089026			089026 mus musculus
7	1455.5	20.0	1651	11	055005			055005 rattus norv
8	1361.5	18.7	1273	5	044928			044928 caenorhabdi
9	1318.5	18.1	1060	11	Q9Q213			Q9q213 rattus norv
10	1256	17.3	1344	11	Q9Z214			Q9z214 mus musculus
11	1230	16.9	232	5	Q9VQ07			Q9vq07 drosophila
12	783.5	10.8	166	5	Q9VQ09			Q9vq09 drosophila
13	717	9.9	423	5	P15172			P15172 caenorhabdi
14	658	9.0	874	5	001632			001632 caenorhabdi
15	603	8.3	1026	11	Q62845			Q62845 rattus norv
16	602	8.3	1028	11	Q62682			Q62682 rattus norv
17	597.5	8.2	1493	11	P97798			P97798 mus musculus
18	595.5	8.2	1427	13	Q91562			Q91562 xenopus lae
19	593	8.2	1232	13	Q90284			Q90284 carassius a

20	592	8.1	1151	11	Q9QVNS			Q9qvns rattus sp.
21	589.5	8.1	1377	11	P97603			P97603 rattus norv
22	581	8.0	1028	11	Q07409			Q07409 mus musculus
23	579.5	8.0	1277	13	Q98902			Q98902 fuqua rubrip
24	578	7.9	1264	5	P91767			P91767 manduca sex
25	576.5	7.9	2016	5	Q9NBAL			Q9nbal drosophila
26	575.5	7.9	1461	4	Q92859			Q92859 homo sapien
27	575.5	7.9	1461	4	000340			000340 homo sapien
28	575.5	7.9	2016	5	Q9V4V9			Q9v4v9 drosophila
29	572	7.9	1217	11	P97685			P97685 rattus norv
30	565.5	7.8	1272	13	Q90924			Q90924 gallus gall
31	565.5	7.8	1445	11	Q63155			Q63155 rattus norv
32	560	7.7	1443	13	Q90610			Q90610 gallus gall
33	553.5	7.6	1180	4	015051			015051 homo sapien
34	552.5	7.6	1280	13	Q90933			Q90933 gallus gall
35	551.5	7.6	1040	13	Q9W675			Q9w675 brachydanic
36	551.5	7.6	1302	5	061542			061542 drosophila
37	547.5	7.5	1166	11	Q9QVNA			Q9qvna rattus sp.
38	545.5	7.5	1154	11	Q9QVNB			Q9qvnb rattus sp.
39	543.5	7.5	1018	6	Q28106			Q28106 bos taurus
40	543.5	7.5	1028	4	Q9Q052			Q9q052 homo sapien
41	543	7.5	1197	13	Q90478			Q90478 brachydanic
42	540.5	7.4	1028	11	P97528			P97528 rattus norv
43	538	7.4	1215	11	P97686			P97686 rattus norv
44	536.5	7.4	1822	4	Q9ULI7			Q9uli7 homo sapien
45	535.5	7.4	1028	11	Q9JMB8			Q9jmb8 mus musculus

ALIGNMENTS

RESULT 1
Q9VQ10
ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
DT 01-MAR-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE C55481 PROTEIN.
GN C55481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydrogata; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celislaier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gale R.F.,
RA George S.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailly L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Perriera S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegwan C.,
RA Jalali M., Kalush P., Karpen G.H., Ye Z., Kennison A.J., Ketchum K.A.,
RA Kimmel B.E., Kohira C.D., Kraft C., Kravitz S., Rulif D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzay D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: A0003586; AAF51373.1; .
 DR HSSP: P56276; 1TLX.
 DR FLYBASE: FBgn0031341; CG5481.
 DR INTERPRO: IPR001412; .
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 1.
 DR PFAM: PF00047; ig; 5.
 DR PRINTS: PR00014; FWTYPE11.
 DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
 SQ SEQUENCE 823 AA; 89715 MW; 36FC0B91P36ZP19 CRC64;

Query Match 53.6%; Score 3895.5; DB 5; Length 823;
 Best Local Similarity 91.7%; Pred. No. 1.4e-266;
 Matches 759; Conservative 0; Mismatches 2; Indels 67; Gaps 3;

QY 11 MDTVFKNQDPTFNQAGCNPTPIQWFKDGRKLTDTGSHRMLPAGGLFLKVIHSSR 70
 DB 1 MDTVFKNQDPTFNQAGCNPTPIQWFKDGRKLTDTGSHRMLPAGGLFLKVIHSSR 60
 QY 71 ESDAGTTCNKAENFGVARSNATLQVAFRLDEFLPEANRTVAGGEVAMCEGAPGSP 130
 DB 61 ESDAGTTCNKAENFGVARSNATLQVAFRLDEFLPEANRTVAGGEVAMCEGAPGSP 120
 QY 131 EPQISNRKNGQITLWGVNKRIRIVDGNMIAQEARSDGRTQCVKVNVTGRESATFL 190
 DB 121 EPQISNRKNGQITLWGVNKRIRIVDGNMIAQEARSDGRTQCVKVNVTGRESATFL 180
 QY 191 KHVVRPFLRGPQNTAVGSSVFCRIGDPLDVLWRRSTGNSMPLRKSFWLHSS 250
 DB 181 KHVVRPFLRGPQNTAVGSSVFCRIGDPLDVLWRRSTGNSMPLRKSFWLHSS 229
 QY 251 GRHVHLEDRSLKLDVLEDMGEYTCADNAVGGITATGLTVHAPKFFIRKNGLVEI 310
 DB 230 GRHVHLEDRSLKLDVLEDMGEYTCADNAVGGITATGLTVHAPKFFIRKNGLVEI 289
 QY 311 GDBVLEFCQNGHPRPLTWSVGNSSLLLPYRGDRMVTLLPBGSRVLSIARFARDS 370
 DB 290 GDBVLEFCQNGHPRPLTWSVGNSSLLLPYRGDRMVTLLPBGSRVLSIARFARDS 349
 QY 371 GKVTYCNALNAGSVSSRTVSVDTQFELPPIIEQGVPQTLVKSIVLPCRTGLTPV 430
 DB 350 GKVTYCNALNAGSVSSRTVSVDTQFELPPIIEQGVPQTLVKSIVLPCRTGLTPV 409
 QY 431 PQVSWYLOGIPIDVQEBERRNLSDAGALTISDQREHDEGLTYCVASNNKSGSSGTLR 490
 DB 410 PQVSWYLOGIPIDVQEBERRNLSDAGALTISDQREHDEGLTYCVASNNKSGSSGTLR 469
 QY 491 LDTPTNKNKFFRAPELSTPGPGKQVYKGENSVLWTRSNKGVSSGLVGYIMF 550
 DB 470 LDTPTNKNKFFRAPELSTPGPGKQVYKGENSVLWTRSNKGVSSGLVGYIMF 529
 QY 551 GKNEDDQVAVGTQVNTTFTQTLGLLPVNYFFILRAENSHGLSPSPMSEPIVGTG 607
 DB 530 GKNEDDQVAVGTQVNTTFTQTLGLLPVNYFFILRAENSHGLSPSPMSEPIVGTG 589
 QY 608 -----RYFNSGLDSEARSLASGCVVLSNASTVDSMKL 644
 |||||||

DB 590 ENESTFLMPFLIRHYFDSLFRPQFNKSCGLDSEARSLASGCVVLSNASTVDSMKL 649
 QY 645 TW-----OIQNKYGEYGVYTAQRLPNPVPVNPAP 674
 |||||||
 DB 650 TWCVNCRDGLSIAAPHSIAHRSLASQMLQINKYGEYGVYTAQRLPNPVPVNPAP 709
 QY 675 VTSNTVPLGISTSTASASASALISTKPNIAAGKRDGETNGSGGAPTLNTKYML 731
 |||||||
 DB 710 VTSNTVPLGISTSTASASASALISTKPNIAAGKRDGETNGSGGAPTLNTKYML 763
 QY 735 TILNGGASSCTITGLVQVTLFEPIVPPYKSVGKPSNRSIARTLED 782
 |||||||
 DB 770 TILNGGASSCTITGLVQVTLFEPIVPPYKSVGKPSNRSIARTLED 817

RESULT 2
 Q9VPE6
 ID Q9VPE6 PRELIMINARY; PRT; 859 AA.
 AC Q9VPE6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE C5423 PROTEIN (FRAGMENT).
 GN C5423.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Hydrozoidea; Drosophilidae; *Drosophila*.
 RN NCBI_TaxID:7227;
 KW [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE:20196008; PubMed:10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Branton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wahl J.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Flannkuch C., Baldwin D.,
 RA Bailew K.M., Basu A., Baxendale J., Bayraktaroglu L., Bensley E.M.,
 RA Benson R.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck Y.J., Brokstein P., Brottler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris M.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegwan C.,
 RA Jalali M., Kailush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Muzay D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: A0003586; AAF51388.1; .
 DR HSSP: P56276; 1TLX.

RESULT 3

```
ID Q9W213 PRELIMINARY; PRT; 1395 AA.
AC Q9W213;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROBO PROTEIN.
```

```
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```

RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.C., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoeh C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Boutler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.

RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ithegam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Zamboni M.P., Zeng H., Zeng L., Zeng Y., Zeng Z., Zeng Z., Zeng Z.,

RA Dasko P., Lei T., Levitsky A.A., Li Z., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.N., Moy-M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Muckern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Polcast V., Prineas J., Sundberg D.P., Suter C., Tachibana T., Thomsen

RA Albert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,
 RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zayori S., Zhao M., Zhang C., Zhao C., Zhang J.

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu B.O.
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., Zhu X., Smith H.O.
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL: AE003458; AAF46887.1; -.
DR HSSP: P56276; 1TLK.

```
DR      FLYBASE; FBgn0005631; robo.  
DR      INTERPRO; IPR001777; -.  
DR      INTERPRO; IPR003006; -.  
DR      PFAM; PF00041; fn3; 3.  
DR      PFAM; PF00047; lg; 5.  
DR      PRINTS; PRO0014; FNTYPEIII.
```

```
SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13F0 CRC64;
```

```
Query Match 24.6%; Score 1790.5; DB 5; Length 1395;
```


Db 114 PFTTITQMKRKGDOGEYWCNKRNRQVGRSHASLQIAVLADDFPEKPTVNAGETAL 173

Qy 121 MEOGAPGSPGPEQISWKNKG-----QTNLVGNKRIRIDVGGNLAQEARQSDGRTQC 174

Db 174 LECGGPPEGPETPTLWIKDGVLDLAKMSGASSRVLVDGGNLLSNVPEIDBENTKC 233

Qy 175 VKNVYQVRESATLAKVHVRFLIRGQNPQAVGSSVFCRIGDPLDVLWRRITS 234

Db 234 IAGNLVGTRESSYALILVQKPYFKPEKQDQVLXQYATFHC SVGGDPFKVLMKK--E 291

Qy 235 GGNMPLKFNLSHSGVRVLEDSKLDDVDLEMGVETCRADNAGVGTATGILTVH 294

Db 292 EGNTPVSRARILH-----DEKSLISNITPTDEGTVCADNAGVGTATGILTVH 342

Qy 295 APPKVFIRPKNQLGIDGVLEQACNGHPRPTLYASVGENSLLLPYEDRGMVETPT 354

343 APPNTPKRSPNKKVSLNGVQLPCMASGNPPSYFWTKEGVSTLMPNSHGRCVTA-- 399

355 EGRSVLSIARFARSGKVTCTNALNAGVSSRTVSVSDTQFELPPIIBQGVNQLP 414

Db 400 -ADGTIQLIDVQDEGTYT--CSAFSVDSSTVRVQLVSVDERPPIIQIGAPQTLP 457

Qy 415 VKSIVLPCRTLGPVQVSKLDGIPIDVQEHERRNLSDAGALISDLARHEDGLTYC 474

Db 458 GSKVATLPCRTAGSPRIKVPDCHA--VQAGNYSIIQSSSLVDQLQ--SDGTYTC 514

Qy 475 VASNRKNGSSVSGSLRLDTPNPNIKFPAPELSTYPPGPKQPNVKEGNSVLSWTRS 534

Db 515 TASGERGETSNAALIVKPEQSTL--BRAADPTYPAPPGQPKVLMVSTISLWAKS 572

Qy 535 MKVGG--SLVQVLEMGKNTDGVNAVCTRWNTTPTQGLLPGVWVFLIARNSHG 592

Db 573 QKXPGAGVPIQYVETPSPALQGMVAAHVGQTVIISGLPTSTVPLVRAENQGS 632

Qy 593 LSLPSPHSPETVGRIFV--SGLLSEARASLSQVLENSVNAVDSLSMLTQW-- 648

Db 633 VTSVPSLSVITKIDDFDASANDLSAAR--LITKGVLELIDASINAVLEMLVH 691

Qy 649 -INKIVEGFTYVQLPNFVWNPAPVTSWNPVLGSTTSASASASALISTFKPIA 707

Db 692 SADEKIVEGLRIHEDASV----- 711

Qy 708 AAGKRDGETNGSGCAPTPTLTKRLMILTGGGASSTITGLVQVTLIEFFVPTFKS 767

Db 712 -SAQTHSITMD--ASAESFVGMKKKITYETFFLIPFFETI 750

768 BGKPNRSIARTLSEVSEAPGEMALLNSAVFLWKAPELKRDRHGVNLHVIVNGI 827

Db 751 EGQSPNSKATLYTSEVSPAPDNIQKMYTAGWVWPPPSQHNKLNKGVKIV-- 807

Qy 828 DTAHNSRILTNVDAASPTLVLANLIGVNTVGVAAAGNAGVGPVCPVATLRDPT 887

Db 808 -SAGNTMVLAMRLNATTSVLLNLTGAVSVRLNSPTTAGGQPSKPLSDPND-T 865

Qy 888 KRLDP-----FINRDH--VNDVLTQWFTILLGAI 916

Db 866 HNVHPFRAPSPGTDGHRGQDLTYHNNGINPPDINPTTHKKTITLDSGLWMLVLCV 925

Qy 917 LAVLMLSG-AMVYFKRHKHMKQ--SALNTMGNHTSDVLMKPLSARNGNGLVDSGTS 974

Db 926 LLVLVISAISIMYFKRKHMTKELGHLGVSDN-----ETALNLSKESLWIDHRHG 979

Qy 975 GMYRVPSPGSDSLMKRDIADYAPVCAQPGSPAGGTSSSGSGGAGSAGSGDDINGH 1034

Db 980 --NR--TADTDD-----SGLSSEKILLSHN 1001

Qy 1035 GSERNQGRVETSNIP--TDTAEVSSFGKAPSEGRHNCASAPATYSILSPHQOQO 1092

Db 1002 SSQSN-----YKNSDGGTDAEVDTRNLITFNCKRSPMDPTATTIMIGTSSSETC 1054

Qy 1093 QQRVYQRPVQVGLQRPMPHPVQOQOQQOQHQH--QALQOQRLPSPNITQOM 1150

Db 1055 TKTTSISAKDS-GRSPSYDAFQAPQVPPVFNKSVNLYQVPEPINWSEFLP----- 1106

Qy 1151 STTSEIYPTNTPGSRVSYSEQYITPKDKQRHIT-----ITENLNSCHTYEAPAGKQS 1204

Db 1107 -----PPEHPPPSTGYTAGQSPESRSKSSKAGSIGSTNQSLINASSHSSSGFSA 1155

Qy 1205 SPISSQFASVBAQLPP-----NCSIGRESARFVLTNDQCKNNLLDLQSSWCY 1256

Db 1160 WGVSPQYVA-----CPENNVISNPLSAVAGTQNRQITPTNQHPQLPAY----- 1206

Qy 1257 NGLADSGCGSSPMMMLSHDEHALYHTA-----DGLDDMERLYVKVDEQPP 1307

Db 1207 --TATGCGGAVPPNLLPATORHAASQOAGLNAARCAQSRACNSCDALTPSQMPPP 1264

Qy 1308 QQQQQLIPLV-----POBAEGHLSQWRNQSSTSSRKNQCEIKPSGLIYAP 1355

Db 1265 P-----VVPFGWQTVPHNPSHMPHTSNHQLY-----QCSSECD--HSR 1304

Qy 1356 GSVASERILLSNSGSSQSSQAGHN 1380

Db 1305 SPSHKKRLQLEEGSSAKQRQGH 1329

RESULT 5

Q9Y6N7

ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.

AC Q9Y6N7;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE RC02NDABOUT 1.

GN ROBO1

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98117249; PubMed=9458045;

RA Kiddle T., Brose K., Mitchell K.J., Patter R.D., Tessier-Lavigne M.,

RA Goodman C.S., Tear G.;

RT "Roundabout controls axon crossing of the CNS midline and defines a

RT novel subfamily of evolutionarily conserved guidance receptors.";

RL Cell 92:205-215(1998).

DR EMBL: AF040990; AAC39575.1; .

DR HSP: P56276; 1TLK.

DR INTERPRO: IPR001777; .

DR INTERPRO: IPR003006; .

DR PFM: PFM00041; fn3; 3.

DR PFM: PFM00047; 1g; 5.

SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 20.6%; Score 1498; DB 4; Length 1651;

Best Local Similarity 28.9%; Pred. No. 9.5e-97;

Matches 410; Conservative 221; Mismatches 526; Indels 260; Gaps 42;

Qy 4 PRIIEHPMOTVTPNDPTFNCAQBNPTPIQWFKGRELKTDG--SHRMLPAGGL 60

Db 58 PRIIEHPMOTVTPNDPTFNCAQBNPTPIQWFKGRELKTDG--SHRMLPAGGL 127

Qy 61 FFLKVIHRS--ESDAGTYCEAKNEFGVARSNATLQVAVLRDEFLPANTRVQAQVEA 119

Db 128 FFLRIVHGKSRPDSGTYVCAVNLGAVSHNASLEVALRDFQPNFSDWVWAGPEA 187

Qy 120 LMWCGAPGSPGPEQISWKNKGQTLMLVGNKRIRI--VDGNNLAQEARQSDGRTQVKN 178

Db 188 WMECQPPHSGPEPISWKNKGSRPLD--DKERITRGGKLMITTKSDAGRTVCVGTN 244

Qy 179 VYQVRESATLAKVHVRFLIRGQNPQAVGSSVFCRIGDPLDVLWRRITSASGN 238

Db 245 WYGERSEVALYVLERPSVVRPSNLAVTVDSAEFCIARGDVPVTVWKRK--DQGL 302

Qy 239 PLRKFSLHSAAGRVHLEDRSLKLDVLEDMGYTCEADNAVGGITATGLTVHAPK 298
 Db 303 P-----KSRTEIRDDHTLKRKYTAGDMGYTCAVNMVWKAASATLTVQEPFH 352

Qy 299 FVIRPKNQLVEIGDEVLEPCQANGRPRLTYVSGNSLL-----PGYRDMGEVTLTPE 355
 Db 353 FVYKPDQVALGRVTYFQCAATGNPQPAIWRREGSQNLFLSYQPPQSSSRFSVSQTGD 412

Qy 356 GRSVLISARFAREDSGKVTCTNALNAVGSVSRITVSV-DTQFELPPPIIBQGPWQTLF 414
 Db 413 ---LTIITVQRSDVGYII-CQTILNAGSIIITKAYLEVDVIADRPVPIVROGPNQWYA 467

Qy 415 VKSIVLPCRTLGTPVQVSWYLDGIPIDVQHERNNLSAGALITSLQREDDEGLTTC 474
 Db 468 VDGTFVLSCVATGSPVPTILWRKGVLYSTQDSRIKQLEN-GVLIQR-YAKLGDTGRYTC 525

Qy 475 VASNRKNGKSSMSGYLRLD-----TPTNPNKIFPAELSTYPPGPGKQWVEGEN 525
 Db 526 IASTPSGEATWSATIEVQEPGVQPPRTPDNL-----IPSPKSPVETVSRN 575

Qy 526 SVTLNWSKNGSGGSLGVYIEMFGKNETDQWAVGTQVMTPTQGLLPGVKNVPLLI 585
 Db 576 VTIVLSN-QPNLSGAPPTSITITAFSHAGSSQWTAZNVKRTETSAIKLAPNATILFL 634

Qy 586 RAENSIGLSLSPNSEPI-TWGTFRNSGLDLSBARSLSDGVWELSNASVDSSTMKL 644
 Db 635 RAANATYISDPSQISDPKTKQDVLPTSGVQHVQVORE-LGNVAHLNPNVLSSSSIV 693

Qy 645 TWQI-INGKYKGFVTVARQLPNPNIWNPAPVTSWNLGSGTSASASASALISTK 702
 Db 694 HWVDQSQSITIGKILYR----- 712

Qy 704 PNIAAAGKROGETWGGGAPPTPLMTKYMILNNGGASSCTITGLVQVITLFFPIVP 763
 Db 713 ---PSGANHGESDMLVFEYRTP-----AKNSVLPDLANGVTEIKARPP 754

Qy 764 YKSVGKPNRSRIARTLEDVSEAPTEGMAILL---NSSAFKLKWKAPELDRGHWLYNTH 821
 Db 755 FNEQSGADSEIFAKTLEAPSPAPPGVTVSKNGDGTALISWGPFPEDTQNGVQETK 814

Qy 822 VIVRGDITAHNFSRLITWIDAASPLVLNLTCEVMTVGVGAAGNAGVGPICVPATL 881
 Db 815 WMLGNETRYRI-----NKTVDGSTVPIPLVPGIRTSVEVAASTGAGSGVSEPOFI 869

Qy 882 RLDPITKRLD---FINGRHNDVNLTPWFIILGAILLAVLSFGAMVYKRRKMMQ 939
 Db 870 QLDAGNPNVPSFDQVSLAQISDVYQPAFIAGACWILMVSIMUL---RRKKRN 926

Qy 940 SALNTWGHNTSDVLMKPSLS-----ARNGNTWLDSTGGMYWPSPGGDSLEWQD 992
 Db 927 GLISTAG---KRPVSPFTPTPTVYQRSGBAV---SSGG-PRGLNLSSEAPAQ 974

Qy 993 HIADYAVCGAGSPAGGTSGGSGGAGSG---ASGGDIDHGGHSENRQVIV--- 1044
 Db 975 WLADTPWNGHNDHDSISCTAGNGSDNLTYSRPACDANTNGLQNLQNTMLNLE 1034

Qy 1045 ---GEYSNIPDTAEVSSPKAPSEYGRHNAS---PAPATSTSLSPHQOQQQQQPRITQ 1098
 Db 1035 STYGVG-OLSNINEMKTFSNPLKDGFRFVPSQPTPATYQLGLNSLNKNN--- 1089

Qy 1099 QRPVPGTGLRPMR---PHYQQOQH---QQOQAQTHQOHALQOQHPSPNITYQMSST 1153
 Db 1090 ---GSGDSGKHNKPLQOQKQEVAPYQVINYQNKLKANDRAANDVYPTPIYNGS--- 1141

Qy 1154 SEIPTNTGSRSVTSBQYTYFKDKGRHITENKLSNCHTETAPAGAKSSPISQFAS 1213
 Db 1142 ---TDNTGGG---TNSSD-----RGSSTGSGSQ--- 1164

Qy 1214 VRRQQLPNCISIGRESARFKVLTDOGRNQNLDDLSGSMCYNGLADSCGGSPSPMAM 1273
 Db 1165 ---HKGARTPKVPKQGGHMDLIL-----PPPAH 1192

Qy 1274 LMSDEHVALHTADGOLD---MERLYVKVDEQPPQOQQQLLPLVQHPABGL 1326

Db 1193 PPFPSNSEYNTVSDVSTQDMPCVPPARMYLOQDELEEDERGPTVPVPGASPA 1252

Qy 1327 QSWNQSSTSSRKNMQE-----CIKPESELIYAP 1355
 Db 1253 VYSHSQSTATLTPSQBELQMLQDCPEETGHMQHP 1289

RESULT 6

OB9026

ID OB9026 PRELIMINARY; PR: 1612 AA.

AC OB9026;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE DUT1L PROTEIN.

GN ROBO1 OR DUT1L.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID:10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RA Wu M.C., Lowe N., Fordham R., Rabbits P.;

RT "The mouse homologue of human DUT1L/H-robo1 gene: protein sequence and

RT chromosomal location."

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; L17793; CAA76850.1; +.

DR HSSP; P56276; 17LA.

DR MGD; MGI:1274781; Robo1.

DR INTERPRO; IPR001777; +.

DR INTERPRO; IPR003006; +.

DR PFM; PF00041; tns3; 3.

DR PFM; PF00047; tns3; 5.

SQ SEQUENCE 1612 AA; 176406 MW; 5F298C8544796AB CRC64;

Query Match 20.5%; Score 1489.5; Db 11; Length 1612;
 Best Local Similarity 28.9%; Pred. No. 3.6e96;
 Matches 402; Conservative 231; Mismatches 516; Indels 243; Gaps 45;

Qy 4 PRIIEHPMOTVTPKNDPFPNCAQNGNPTIQWKGRELKTDG---SHRIMLPAGGL 60

Db 29 PRIIEHPMOTVTPKNDPFPNCAQNGNPTIQWKGRELKTDG---SHRIMLPAGGL 60

Qy 61 FFLKVIHNSR-ESDAGTICEAKNFGVARSNATQIQAVALRDGFLEPLRNTAVQAQVGA 119

Db 89 FFLRITVGHKSRPDGCVITCAVARYLGEAVSHASLEVALRDGFQNPQSDVWVAQVGA 148

Qy 120 LMBGAPRGSPEPQISWRKNGQTLNWKGRIRI-VDOGNLAIQBARQSDGRTQCVKYN 178

Db 149 VMBQPPHGPPEPTISWKKDGSPLD---DKDERTIGKLMITTKRSAGKTYVCVGN 205

Qy 199 VYGTRESATLKYVHFRPLIPQNPQATVAGVSSVQTCRIGDPLVPLWNRATASGNN 238

Db 206 MYGERSSEVATLVLERSPFVRRGNLAVVDGAEFKCEARGDVPTVYWRK---DDEL 263

Qy 239 PLRKFSLHSAAGRVHLEDRSLKLDVLEDMGYTCEADNAVGGITATGLTVHAPK 298

Db 264 P-----KSRTEIRDDHTLKRKYTAGDMGYTCAVNMVWKAASATLTVQEPFH 313

Qy 299 FVIRPKNQLVEIGDEVLEPCQANGRPRLTYVSGNSLL-----PGYRDMGEVTLTPE 355

Db 314 FVYKPDQVALGRVTYFQCAATGNPQPAIWRREGSQNLFLSYQPPQSSSRFSVSQTGD 373

Qy 356 GRSVLISARFAREDSGKVTCTNALNAVGSVSRITVSV-DTQFELPPPIIBQGPWQTLF 414

Db 374 ---LTIITVQRSDVGYII-CQTILNAGSIIITKAYLEVDVIADRPVPIVROGPNQWYA 428

Qy 415 VKSIVLPCRTLGTPVQVSWYLDGIPIDVQHERNNLSAGALITSLQREDDEGLTTC 474

Db 429 VDGTFVLSCVATGSPAPPTILWRKGVLYSTQDSRIKQLEN-ESGLIQR-YAKLGDTGRYTC 486

```

DS TRANSMEMBRANE RECEPTOR RB01.
OS Rattus norvegicus [Rat].
OC Karyota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10115;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
EX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cite 92:205-215(1998).
DR EMBL: AF041082; AAC39960.1; -.
DR HSP: P56276; 1TLK.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; Ig; 5.
KW Transmembrane.
SQ SEQUENCE 1651 AA; 180746 MW; FA2452DDA4E18687 CRC64;

```

Query Match : 20.0%; Score 1455.5; DB 11; Length 1651;
Best Local Similarity 28.8%; Req. No. 9.5e-94;
Matches 406; Conservative 229; Mismatches 525; Indels 251; Gaps 45;

```

QY      4 PRRIEHPCTTVPKNDPTFFVQAOAGNPTTIFQWKGDEGLTDTG---SERIMPAAGL 60
Db      68 PRVIEHPDGLVMSKGPATLNCKAGAPPTTIEWYKGERVETDKDDPSRMLPSGSL 127

QY      61 FFLVLHHSR-ESDAGTCEAKNEFVARSRNATQVAVLEDFEFLPANTVQAGEVA 119
Db      128 FFLRLVHSGRSRPGDSVTCVARNYLVGEVSHASLEWALRDQFQNSDVVAVGEPA 187

QY      120 LMECAPPGSPPEPISWRNKQNTLMLVGNKRRI-VDGGNLAIEARQSDGRQVYKN 178
Db      186 VMQCQPPGHPPEPTISMKDGSPLD---DKDERTIRGGKMLTITRKSDAGKTVCGTN 244

QY      179 VGTRESATATLKHVHPYFLRIGQNPQTAVGSSVYFGCRIGDPLPOVLWRRTSAGNM 238
Db      245 VHGERSKDVVLTLEPFSRVPKSLNVLVDDSAEFCEARGDPPTFGWM---DDGEL 302

QY      239 PLAKFSWLSASGRVWLEDRSLKDDVTLDEMGEYTCADNAVGITATGLTVHAPPK 298
Db      303 P-----KSYRELDHDLTKRKVATAGMGSCTVADNVMGKAEASATLVQEPHH 352

QY      299 FVIRPKMQLVEIGDEVLFEQAGHPRTPLTWSVEGNSLL---PGTRGRMVTLTPE 355
Db      353 FVVPKRDQVAGLVATPTQCBAATGNOPALVNRKESQNLFLSQTPQSSSRFSVGTGQ 411

QY      356 GRSVLIAERFAERDSKVTVCNALNAGVSRSRTVSVY-DQFELPPPIEQGVNITLP 444
Db      413 ---LTVNVSRSVGTYT-COTLNAVGSITITAYLETVDIADRPPIVIRIGQVNTA 467

QY      415 VKSVILPCRLTGPVQVSVYDLGIDPIDVQEHERRNLSDAGALITSDQRHEDGLTQ 474
Db      468 VDGTLTSLCVATGSPPTILWRKDGVLTVSDQRSLQ-ESGVQLIR-TAKIGDGTGYT 525

QY      475 VASNRNGKSSAGSLRLD-----TPPNINIKFIRAPFLSTPTGPKQPMKVEKEN 525
Db      526 TASTPSGATNSATIEVQEGVPVOPPRPTDNL-----IPASGRPEVDTYSK 575

QY      526 SVTLSTRNSKNGVGSGLVGTVMFNGKNETDGWAVGTQVNTTPTQGLLGNVNFYL 585
Db      576 VTLILN-OPNLSNAGTPTTYIEAFSHAGSSQTVANVNTETFAIKGLPNATILFLV 634

QY      586 RAENSRLGLSLPSPMEI-TWTRFVNSGLDISEAASNLISGDW-ELSNASVDSTSMK 643
Db      635 RAANAGISDPSQISDPTVDQPTPTQGVDRHQVREL--GNVLHLNPTILSSSSVE 692

QY      644 LTWQI-IMGVYEGVFIYARQLPNIPVNPAPVSTNPLIGSTSSASASASALIST 702

```

```

RESULT 7
055005
ID 055005 PRELIMINARY; PRT; 1651 AA.
AC 055005;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

```

Db 693 VHWVDQSQSYTGCTKLYR----- 712

Qy 703 KPNIAAGKRDGETYNGSGGAPPLNTKYRMLTLNGSGASSCTITGLVQTLTEFFIVP 762

Db 713 ----PSGASHGSEMLVEFVETP--TK-----NSVVPDLRKGWVYETKARP 753

Qy 763 FTKSVGKFSKNSRIARTLDVPSAPYGMALL--NSSAVFLKAWPELDRGVLLY 820

Db 754 FFWPEQGADESEKPAKTLERPSAPPSTVYKNDGNGCTAILVWQPPPTQWGMQWQY 813

Qy 821 YHVIVGIDTAHNSRILNTVDAASPTLVLANLTGVMYTVGVAAGNAGVCPYAT 880

Db 814 YVWGLGNETRIH-----NXTVDGSTPSVPIPLPGIRYSEVVAASGAGPCKVSEPO 868

Qy 881 LRLDPLTKRLDGF--FINQRDEVDVLTQWFIILLGALLAVMLSPGMYVYKRRMOK 938

Db 869 IQLDSHGNPVSPEXQVSLAQISDVYKQPAFIAGIGACWILMLWPSLWYRHRK----K 924

Qy 939 QSANTLGMNHNHSDVLMKPSL-----ARNGKTYWLDSTGMYWPSPGGDSLEMKQ 991

Db 925 RNLGLST----YAGIRKYPSTPTPTVYQGGBAV----SSGG----RPGLLNISEPATQ 973

Qy 992 DHIADTAPVCGAPSGAGGTSGGSGCAGSG----ASGGDIBHGGHSGNQRTV---- 1044

Db 974 PHLADTAPVCGHNDSCINCCASNGSDNLTYSRPADCAIYNNQGLDNQTLMLP 1033

Qy 1045 ----GETSNIPDIAEVSFGKAPSETGRGNAS--PAPATSSILSPQQQQQQQPRY 1097

Db 1034 ESTYVYG--DLSNKNEMNTFNSPKLDGPFVNPQGQPTPIATTLQANLNNMN---- 1089

Qy 1098 QQRVPYGTGLRPMHPYVQQQHQQQQAQTHQHQALQHQQQLPSPNITYQNSTSEIY 1157

Db 1090 ----GGG-----DSSKHKPPGQQ--KQEVAPITYNMEQKMLNKDITRANDTIL 1133

Qy 1158 PTNPGTSRVSYSQITTYPKDRKHITHENKLSNCHTYEAPAGCAQSSIPASVYRQ 1217

Db 1134 PT-----IPYN--HSTQNTGGSYNS--SDSGSTSGS 1162

Qy 1218 QLPNCSIG--RESARFYKLTMDQGNQKNDLDDGSMCYNGLADSGCGSPSMXMLS 1276

Db 1163 Q-----CHKGARTPKAPQGGMNWALL-----PPPPAPPP 1195

Qy 1277 HEDHRYALTYADGDLDD-----MERLYFVDEQPPQQQQQLIPVQPAHGLQSN 1329

Db 1196 HSNSETSMSVDESTQRMPCPVPARMYLQDELEEBEARGPTPVVGAASSPAVSY 1255

Qy 1330 RNQSTRSSKNQGCBCIK-----PSELIYAP 1355

Db 1256 SHQSTATLTPSPQELQPMQLQDCPEDLHMP 1286

RESULT 8

044928

ID 044928 PRELIMINARY; PRT; 1273 AA.

AC 044928;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE SAX-3.

GN SAX-3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderiinae; Caenorhabditis.

OX NCBI_TaxID:6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98117250; PubMed=9458046;

RA Zallen J.A., Yi B.A., Bargmann C.I.;

RT "The conserved immunoglobulin superfamily member SAX-3/Robo directs

RT multiple aspects of axon guidance in C. elegans.";

RL Cell 92:217-227(1998).

DR EMBL; AF041053; AAC38648.1; .

DR HSSP; P56276; 17LK.

DR INTERPRO; IPR001777; .

DR INTERPRO; IPR003006; .

DR PFAM; PF00041; fn3.3.

DR PFAM; PF00047; ig.5.

DR PRINTS; PR00014; FNTYPEII.

SQ SEQUENCE 1273; AA; 139427 MW; 0138766B51A7BAD7 CRC64;

Query Match 18.7%; Score 1361.5; DB 5; Length 1273;
 Best Local Similarity 28.4%; Prod. No. 2.8e-87;
 Matches 357; Conservative 202; Mismatches 437; Indels 259; Gaps 37;

Qy 4 PRIIEHPMOTVPKNDPPTFCNQAGNPTPTIQWKGDEL---KTQDSHRMLPAGGL 60

Db 31 PVTHIEPINDVYSGSPATLNCARPSTAKITVWDGQPVITMKQVNSHRVLDQSL 89

Qy 61 PFLKVIHSR--RESADGYWCBKMEFGVARSNATQVAVLRDFEFLPANTHVAQGV 118

Db 90 PFLKVGKNGKSDAGATCYASMEHEGVKNSLKLALHEDFVRPRTVQALGEM 149

Qy 119 ALMECHGSPGPEPQISWKNQQLNLVGNKRIIVDGGNATQIBAQSGSDGQYCVKN 178

Db 150 AVLECSPPAGPFPVVSWRKDELRIQDKPRTILSDGNLIDVDRSDGTCQVANN 209

Qy 179 VVOTRESAPALFKVHVRPFLIRGQNQAVGSSVYVQCRIGDPLDVRMLRSTASGN 238

Db 210 MVGHVSPNAPLSVFEKPAKFBQKMDVYGAVALFDGCVTDGQVQITWKR--KNEPM 267

Qy 239 PLAKFVNLASGRVHLED-RSLAKLDVTEMGYEITCADNAVGITATILTYHAPP 297

Db 268 PVT-----BATIAKDWGLRIERQVSDGEGVETATARNAGTLEASAGLQVAPP 317

Qy 298 KFVIRPKNQLVEIGDEFLPCQANGHPRTILVSVGNSLLAPCT--RQGMVETLPE 355

Db 318 SFQYPAQVSGPAGTATPECTLVGQSPATFVSKSGQQLFVSTYSADGKHYSTP-- 375

Qy 356 GRSLVSLAFKAREDESGKVTCNALNAVGVSSRTYVSDTQF-----ELPMPPIEGQV 409

Db 376 --GLTIEVHQVDEGAY-CAGMNSAGSSLSKALVKTAVTKVNTPAKPTPIEHQ 432

Qy 410 NQTLPVKSVIPLPCLRTLPVQPSWYLGDPIDGVHERRNLSDAGALISQLQHEDE 469

Db 433 NQTLWGSALLPCQASGKPTGSLWROGLDITD-SRLSQSTGSLHIAOLAK-PDT 490

Qy 470 GLITCYASNKNGSSWSGSLRLDTNPNNKIFRABELSTYGPQGPQWGVENGSVTL 529

Db 491 GYVICTAKNDEGSTWSASLTVEDHST-NAQFVMPDPNPNSSPTQPIVNVDTTEVL 549

Qy 530 SWTNSKNGVSSLVGYVEMFGKNETDGVWAGVTRQNTPTTQGLLPGWNYFLIRAE 589

Db 550 HNAWAPSTGAGPTIGIITYSPDLQWENIPDYVASTYRIKGLGSPHSTMYIRAE 609

Qy 590 SHGLSLSPMSSEPTVGYTRFVNSGL-----DLSEARSLSDGVNLSASVSDTSNK 643

Db 610 EKQIGTSPYSVALVTSKPAQVAVLSDKNMDMAIAEKRLTSEQILDEKVTINSTIAVR 659

Qy 644 LTQWQIN-GLAVEGYVYARQLPNPVIYNNPAPVTSWNTNPLGTSSTASASASALIST 702

Db 670 LPWKKRKLLELDGYTIKMR-----GPRTDQGVYVWVSPS----- 706

Qy 703 KPNIAAGKRDGETYNGSGGAPPLNTKYRMLTLNGSGASSCTITGLVQTLTEFFIVP 762

Db 707 -----TENYVSNLMPPTNTEFFIVP 727

Qy 763 FYK----SVGGEKFSKNSRIARTLDVPSAPYGMALLNSAVFLKAWPELDRGVLLN 813

Db 728 YHGVSGISGASNSMVLTAADAPSLPPEVDRIKMLNLTLLRISWAPADGIGMLAG 787

Qy 820 YHVIVGIDTAHNSRILNTVDAASPTLVLANLTGVMYTVGVAAGNAGVCPY-CY 877

Db 788 QFQIVG--QAPNNR---NITTNERASVTLPHVTGMYTKIRVAASNGGVSGTSGTS 842

Qy 878 PATRLRDPITKRL-----DPFIN---QRDEHNDVLTQPFILLGAILAVLMSFGAMV 928
 Db 843 EVINMQDTLXHLAAQENESFLYGLINKSEVP-----VIVIVALLIFVILIAIC 894
 Qy 929 FVK-----RKHMMKQSAINTMRGNTSDVLKMS-----LSARNGTGLV 969
 Db 895 TWNRNSNDGKDRSFIKINDGSVH-MASNNLWDVACPNQNPMTNTAGRTMNNRQAL 953
 Qy 970 DSST-----GGMWRPSPGGDSLEMKDHADIAPVCGAGSPAGGTSGG 1016
 Db 954 TSLTMAQDFTNCCDTSMTMRPQ-----SEHHYIAQLTGGGP----- 994
 Qy 1017 SGAGSGASGGDDHGGHGSERNQRTVGEYSNIPDIAYVSFGKAPSEIGRHGNASPA 1076
 Db 995-----AMSTF-----TGMQYHDPS 1009
 Qy 1077 PYATSSILSPHQ-----QQQQQPRYQRPVPGYGLQRMHP-----HYQQQHQQQQA 1125
 Db 1010 PYATTTLVLNSQQPAMWLNDKMLRAPAMPNTVPV-----PEPPARYADHTAGRSRSRA 1063
 Qy 1126 QQ-----THQHQALQHQQLPPSMI-YQOMSTTSEIYPTNTGSPRSVSQD 1171
 Db 1064 SDGERTINGLHRTSGQRSDSPHTDVTYQLHSSD-----GTGSKERTGER 1113

RESULT 9

Q90213

ID Q90213 PRELIMINARY; PRT; 1060 AA.
 AC Q90213;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxId:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:99200391; PubMed:10102268;
 RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
 RA Tessier-Lavigne M., Kidd T.;
 RT "Slit proteins bind Robo receptors and have an evolutionarily
 conserved role in repulsive axon guidance.";
 RL Cell 96:795-806(1999).
 DR EMBL: AF182037; AAF04558.1; .
 DR HSP: P56276; IRLX.
 DR INTERPRO: IPR001547; .
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; Ig; 5.
 DR PRINTS: PR00014; FNTYPIII.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Receptor.
 FT NOW_TER 1060 <1060
 SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C1B8542D4A CRC64;

Query Match 18.1%; Score 1318.5; DB 11; Length 1060;
 Best Local Similarity 30.84; Pred. No. 2.3e-84;
 Matches 356; Conservative 180; Mismatches 427; Indels 191; Gaps 39;

Qy 4 PRIIEHPMOTVTPKNDPFTNCAQEGNFTPI-----QWFKQRELTDTGSHRM 54
 Db 31 FKXVEQPSVIVSKGKPNTPNWKQGRFPPTIGKQVMKPKQDKTKDSSKVTQG---CL 87
 Qy 55 LPAGLFFLKVHRSR-ESDAGTYCEAKNRRNFGVARSNATQIVLNRFRLEPANTRV 113
 Db 88 LPSSGTLFLLIYHGRSRKPDGDTVCVARNYLGEAVSRNASLEALLRDRFRQNPDMV 147
 Qy 114 AGEVLAEMCGAPGSPGEPQISWRKNGQTLMLVGNKRIRI-VGGNLAIQBARQSDGRY 172
 Db 114 AGEVLAEMCGAPGSPGEPQISWRKNGQTLMLVGNKRIRI-VGGNLAIQBARQSDGRY 172

Db 148 AAGEPAILECQPPRHPPTIYWKXKVRID---EKEERISIRGKLMISNTRSDAGMY 204
 Qy 173 QCVKNNVGTRESATFLVHVRPELIRGQNPQAVGSSVVFQCRIGDGLPVLWRRIT 232
 Db 205 TCVGTNMGESDPAELIVFERPILRRPINQVJLEDEFAEFQVQDQDPQIVYMKK- 263
 Qy 233 ASGNNMLKPSFSLWSASGRVHVEDSLKDDVTLEDGSETCEADNAVGGITATIGLT 292
 Db 264 -DDADP-----RGRTDKDDTLAIKKAISDGTTCIAENSVGVKVESATLT 312
 Qy 293 VHAPKVFYKPNOLVIGDEVLEPCQANGHPRPTLYSVEGNSILLPT-ROGMENVY 351
 Db 313 VRAPPQFVVRPDQIVAGQRTVTPCEITKGNQPAVFWKQSGNLLFPQDPQNSRCS 372
 Qy 352 LTPGERSVLSIARFAREDSGKVTCNALNVGSSRTVSV-DTQFELPPPIIEQGVN 410
 Db 373 VSPGTD-LITNTIQRSDAGTYI-CQALVAGSILAKAQLVETWLDTRPPPIILQSPIN 429
 Qy 411 QTLPSVTVLPCZLTGPPVQSWLGDIPQVQHERM.SDAGALTILQGHDEBG 470
 Db 430 QTLAVDTALLKACQK-PPLVSWLKGFTF-LGRDPRAITQDQGLIKML-RISDTG 486
 Qy 471 LYTICVAMNKNKSGSCYLRLDTPNINIKYFPAPELSTYPCGPKQWKEGNSVLS 530
 Db 487 TTYCVATSSSGETSWSANLDV---TESGATISKNYDNDLPGPSKQVDTKNSVLS 543
 Qy 531 WTPSKNGVSSLVGVYIEMFKNEDGQWAVGRVQNTTITQGLLPGVNYFLIRAES 590
 Db 544 W-QTGPQVPLASATITAEFSQSVNSQVAVNVHTLTLYVLEGRNPTLYLWRAINP 602
 Qy 591 HGLSPSSPSEPT-TWOTRYKGLDLSASARLSDGV-VELNASVDSMTSLWQI 648
 Db 603 QGLSDSPSPSSVPTQDISPAGQVDRQVKELE-GDVTYVLANPLVPTTVPYVTV 660
 Qy 649 -INGVIEGVFYVARLPVFNWNPVTSNPLNGSTTSASASASALISTKPNIA 707
 Db 661 DRQPTQISTRTVTK-----TSLQAST-----VWNLDAKPTPERSAV 700
 Qy 708 AAKGRDGTWVGSGGAPPLNKTMLTNGGASCTITGLQVTLTFEFTVPTKSV 757
 Db 701 LVNLKGVY-----YEIKVPTFNEF 721
 Qy 768 BGKPSNRSNIATLEDVPSAPPGMEALL---NSSAVFKWPAKELKRGVLLAHYIV 824
 Db 722 QGMSSEKTYITTEAPSAPOGVTVLVGSNHTSISVSKGPPADHQIQUETIWC 781
 Qy 825 RGDIANFSRATLIVNIDAASPTVLNLTGVEVMYTVGAAGNNGVPTCPVATRLD 884
 Db 782 LG-----NETRFINKTVDATISVGLGPPQITRVVAATSAGVSKVSEPPITIG 836
 Qy 885 ----PIYKRLDPINQRHNDVLTQPFILLGAILAVLMSFGAMVFKRRHM-KK 938
 Db 837 GRNEVITENNNSITEQ---ITDVQKQPAFAGTAGACVILMGFSPILWTRKKRGLS 893
 Qy 939 QSAINTMNGNTSDVLKMSLARSNGNYLSDSTGGMWVRPS-PG-----GDSLEMKDR 993
 Db 894 NYAVTFQGD-----GGLMSNGSRGLLNGDP---NTPM 925
 Qy 994 IADTAPGAPGSPAGGTSGGSGGAGSGAGSGDDI---HGGHGSERNQRTVGE-TSN 1049
 Db 926 LADSWPATSFLPVNNSGSPNEIGNPGRG-----DVLPPVGGQKATMLSDGATISS 978
 Qy 1050 I-----PIDAYSVSSFGKAPSEYGRHGNASPAFYATSSIL---SPHQ-----QQQQQPR 1096
 Db 979 IDFTTITNTSSSQITQA-----TPYATTLHSNSIHELAVDLPQGMKSSV 1026
 Qy 1097 YQRPVPGTGLQRP 1110
 Db 1027 QQRSDLMFATSLP 1040

RESULT 10

Q92214

ID Q92214 PRELIMINARY; PRT; 1344 AA.
 AC Q92214;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE RIG-1 PROTEIN.
 GN RIG1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.:
 RA Yuan S.-S.F., Cox B.A., Dasika G.K., Lee E.Y.-H.P.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF060570; AB011628.1; -.
 DR HSSP; P56276; ITR1.
 DR WGD; MG1:1343102; b0g1.
 DR INTERPRO; IPR001777; -.
 DR INTERPRO; IPR003007; -.
 DR PFAM; PF00041; fn3.3.
 DR PFAM; PF00047; Ig_C5.
 Q SEQUENCE 1344 AA; 143439 MW; 8B0060341C49CEA CRC64;

Query Match 17.3%; Score 1256; DB 11; Length 1344;
 Best Local Similarity: 27.2%; Pred. No. 8.5e-80;
 Matches 378; Conservative 194; Mismatches 462; Indels 356; Gaps 48;

QY 4 PRIHEHMOITYKANDPFTFWCQAGNPTPIQWFKDRELK---DTGSHRLMAGL 50
 DB 42 PRIVEGPPOLVSGEPATLCRAGRPAPINIKWKGARVATAREDPRAHRLMAGL 101
 QY 61 PFLKVIHRSR-ESDAGTWCENEFKGVARSNATQVALVROEFLPEAPNTVAQGEVA 119
 DB 102 PFERIVHGRSRSPDGYTCVARNYLGAAASRNASLEVALRDQSPGCVNVAQGEVA 161
 QY 120 LMGCGAPRSPPEPQTSWRKNGQTLNLVGNKRIRVDGNLAIQEARQSDGRTQCVKNY 179
 DB 162 WMDPKVPGKHPFLVTKWGK-KIRLKEEGRTIRAGLKMHSHTFYSDAGMTCVASN 219
 QY 180 VGTRESATFLKVIHVRPFLIRPQNTAVVSGSVPCRIIGDPLVLRRTASGNGMP 239
 DB 220 AGERESGAELVLERPSFLRRPQVVLADAPVNLCEVQDGPQLMHRK-DGDEL 277
 QY 240 LRKFSNLSHSGSRVGVLEDRSLKLDVLEDMGRTCEADNAVGGITATGILTVAPRPF 299
 DB 278 -----AGREIHDRSLWIDQVSEDEGTYTCVANSVGRASGSLVAVPPQF 327
 QY 300 VIRKPKQLVEIDGDEVLEFCQANGHPRLTVSWVGNSLLPQYRGRMWTILTPGRSV 359
 DB 328 VTKPQTVAPGANGVSCETKGNPPAIPKQKSGVQLFPSQ-----SLQPMRL 380
 QY 360 -----LSIAKFAREDGRVYTCNALNAVGSVSRTYVSV-DIQELPPPIEQGVNQT 412
 DB 381 VSPRQGLITEVIGGGTYTV-CQAVSVAGSILAKALLEKINGASIDGLPILLAGPANT 439
 QY 413 LPKSVIVPLCRTLGTLPVQVSV-----YLDGIPDQVHERNNKSDAGALISDLQHE 467
 DB 440 LVLSGSVWLCRVIGNPQNIQWKRDERWLGQ-----DSDGFMMDGQHLASIQ-EM 492
 QY 468 DGLGLYTCVANSRNGKSSNGYLRLDT-----PTNINIKFPAPELSTYPPGPK 516
 DB 493 DMGFTYCAKSSIGATWNSLWKQEDGASGPATGSPN-----PGPSS 539
 QY 517 PQWKEGNSVTLNWSRNVKQSSLVGYGVIEFGKNETDGAWAGTVGNTFTTQGL 576
 DB 540 PIVTYVANSILATW-KPMSGATATSVIAPASQAAGNTWRTVAGVQLETITISGLQ 598
 QY 577 PGVNTFFLIRNAGHSLSLSPMSPEITVGTIRVNSGLDSEARSL-----LSDG 628
 DB 599 PNTIYFLVIRAVGANGLSERPSPVPEVQT-----QDSSLRPAEDPWKQGRGAEVA 650

QY 629 VELSNASVYDSTSMKLTWQLNG-KYVGSPFYAR-----QLNPVNVNPAF 674
 DB 651 VMEGPEVTPGRLTLQVSN-TVDGPGVLQVQGFVRSWRIAGLDQGSWIMLDQSP--HQGST 707
 QY 675 VTSNTWLLGSGTSTASASASASALISXPKNIAAAGKROGETNGSGGGATPLNTKRL 734
 DB 708 VLRLGFP-----GAQIQKVVQ-----QQGGLGASPFVTR----739
 QY 735 TILNGGASGCTITGLVLYTEFFIVPFYKSGVSGKPSRIARTLEDPVSEAPYMEAL 794
 DB 740 -----SIP-----TEAPSGPPQGVAVA 756
 QY 795 L-LNNSAVFLNKAPELKRHGVLLNVIVYIGDITAHNFSRLITWITDAISPVLIA 852
 DB 757 LGGDNRSSVTVSMPEPLPSQRNGVITEQICWLG-----NESRFLHARSAGWARSVTS 811
 QY 853 NLZEGVMTVGVAGNNGNAGVGPVCPATLRD-PITKRDLDPFPIQ--RDEYNDVLTPWF 909
 DB 812 GLLPQITRYALVAATAGVGVASAPVLVQALPPPAERPGVESEGLAERLAKVLRPAF 871
 QY 910 IILLGAILAVLMLSPGMVTVKRRKQWQKQALNTMGNHNTSDVLMPSLSARNNGY--967
 DB 872 LAGSSNACGALLGCPAALYRQK---QRKELS---HYTASFTTPAVSFPHSELGS 923
 QY 968 -----WLDSTGQVNRVSPGSGDLEMKDIAADYAVNCGAGSPAGGT 1012
 DB 924 SSSRPMLGCPAATVWLADS-----WPHRPSFSAEPR-----GSCCPNPODDRYT 972
 QY 1013 SSGG-----GDIHGHGSRNQRVQVGEYSNITPYA 1055
 DB 973 NEAGISLILQATARGANSGSPVYSTDPVGEELQTFYGG-----FPQHSQDPSWNS 1026
 QY 1056 EVSSFGKAPSHY-----GRHG-----NASPATYATSSILSPHQQQO 1092
 DB 1027 QT-----APPENSEGDSGARGQKLLKVPQNPISLWPEALPPPPSCSLCPGSEEE 1081
 QY 1093 QQ-----PRYQRPVPGYGLQRMPHPTQQOQR 1120
 DB 1082 LKSSDLEWCPVPPEKSHLVGSSSGACNVAPAPDPTPSTSTG-----1127
 QY 1121 QQQQAQTHQHQALQHQHQLPPSHY-YQOMSTTSEIPTYNTPSRVSYSEQTYTPKOK 1178
 DB 1128 QGSTATLTPSPDPDPQ-----PTDIPHLMPRRVPL-----GQSPSLVSPALSHD 1177
 QY 1179 QRRIHTENKLSNCH-----TYEAPG-----AKQSPISSPASVPEQ-----1217
 DB 1178 GRPVLGAGPVLSTHASPSPVPSTASSAPGRTRQVGTGHTPLRGRHRIKXKPAALPR 1237
 QY 1218 -----QLPP 1221
 DB 1238 RHNSGDLPL 1247

RESULT 11

ID Q9VQ07 PRELIMINARY; PRT; 232 AA.
 AC Q9VQ07;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE C05574 PROTEIN.
 GN C05574.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravits S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AB003586; AF51376.1; -
 DR FLTBASE: Fp0001338; CG5574.
 SQ SEQUENCE 232 AA; 25580 MW; 8EB530901DECAEDA CRC64;

Query Match 16.94; Score 1230; DB 5; Length 232;
 Best Local Similarity 100.00; Pred. No. 3.8e-79;
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1150 MSTSTSTPTPTGPGSRSVSYSDYITTPDKQRIHITENKLSNCHTYEAPGAKQSPSS 1209
 DB 1 MSTSTSTPTPTGPGSRSVSYSDYITTPDKQRIHITENKLSNCHTYEAPGAKQSPSS 60

QY 1210 QFASVVRQQLPWCISIGRSARFVLTQDKNQMLLDGSSGCTMGLADSGGSGPS 1269
 DB 61 QFASVVRQQLPWCISIGRSARFVLTQDKNQMLLDGSSGCTMGLADSGGSGPS 120

QY 1270 PMAHMSHDEHALHTADGDLMDERLVKVDQPPQOQQQLILPVPQPHAGLQSN 1329
 DB 121 PMAHMSHDEHALHTADGDLMDERLVKVDQPPQOQQQLILPVPQPHAGLQSN 180

QY 1330 RMOSTASSRKNQBCIKRPSSELYAPGSVASERSLLNSGGSTSQPAGBNV 1381
 DB 161 RMOSTASSRKNQBCIKRPSSELYAPGSVASERSLLNSGGSTSQPAGBNV 232

RESULT 12
 Q9VQ09
 ID Q9VQ09 PRELIMINARY; PRT; 166 AA.
 AC Q9VQ09;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE CG14348 PROTEIN.
 GN CG14348.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE:20196006; PubMed:10731132;
 RA Adams D.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gale R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravits S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AB003586; AF51374.1; -
 DR FLTBASE: Fp0001340; CG14348.
 DR INTERPRO: IPR001777; -
 DR PFM: PFM0041; fn3; 1.
 SQ SEQUENCE 166 AA; 18353 MW; 5FDFD163A17C217 CRC64;

Query Match 10.88; Score 783.5; DB 5; Length 166;
 Best Local Similarity 96.94; Pred. No. 7.8e-48;
 Matches 156; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 791 MEALLNSAVFLKWKAPLKDHRGVLLNVHVIIRGIDTANFRSLITNVITDAASPLV 850
 DB 1 MEALLNSAVFLKWKAPLKDHRGVLLNVHVIIRGIDTANFRSLITNVITDAASPLV 60

QY 851 LANLIEGVHTVTVGAAGNAGVGPVCPATLRLDPITKLOPFIINR---DHVNDVLT 905
 DB 61 LANLIEGVHTVTVGAAGNAGVGPVCPATLRLDPITKLOPFIINRPTINQDHDVNDVLT 120

QY 906 QPWFIITGAILLAVLMSFGANVFKRKNMNSQALNTMR 946
 DB 121 QPWFIITGAILLAVLMSFGANVFKRKNMNSQALNTMR 161

RESULT 13
 P91572
 ID P91572 PRELIMINARY; PRT; 423 AA.
 AC P91572;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
 GN ZK377.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproot J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Nhan M., Hawkins J.;
 RL Submitted (FEB-1997) to the ENBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1997) to the ENBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the ENBL/GenBank/DBJ databases.
 DR ENBL: U98183; AAB52658.1; -.
 DR HSP: P56276; ITLX.
 DR INTERPRO: IPR003006; -.
 DR PFAM: PF00047; Ig; 4.
 SQ SEQUENCE 423 AA; 4654 MW; B4530DB6D575B5 CRC64;

Query Match 9.9%; Score 717; DB 5; Length 423;
 Best Local Similarity 39.18; Pred. No. 1.7e-42;
 Matches 155; Conservative 61; Mismatches 154; Indels 26; Gaps 9;

QY 4 PRIIEPMQTVFKNPPFTNCQAGNPPPIQWKGDEL---KIDTGSHRMLPAGGL 60
 DB 30 PVIIEHPIDVWVSGSPATLNCAGKPS-TAKITWYKDGQVITNKEQVNSHRVLDTGSL 88
 QY 61 PFLKVIHSR--RESGDAYTCWAKNEFGVARSNATLQVAVLDEBFLEPANTVQAQVE 118
 DB 89 PFLKVNKGKNGKSDAGATYCVASNEGEVNSNEGSLKMLMLEDPGVRPRFTVQALGGEM 148
 QY 119 ALMECAGPSPPEQISNKRNGQITLNLVGNKRIRIVDGNLAIQEARQSDGRTQYCVKN 178
 DB 149 AVLKCSPPPPGPEPVSNRKKDELRIQMDPRYLSGDNLIDVDRSDSGTQCVAIN 208
 QY 179 VVGTSASATFLKHVHRFLINGPONQZAVVGSVVPQCRIGGDLPOVLNKRRTASGGNM 238
 DB 209 MGVERVSNPARLSVFERPKFQBPQKMDTVGVAVLPDCRVGDPQIQITWKR--ENEM 266
 QY 239 PLRKFSLWLSASGRVHVED-RSLKDDVTLEMGYTCADNAVGGITATGILTHAPP 297
 DB 267 PVT-----RAYIAKDNGLRIERVQPSDEGVTCYANPAGTLEASALRLVQAPP 316
 QY 298 KFYIRPNQVLVEIGDEVFQCAQNHPRPTLVNYSVGNSSLLPQT--ROGRMEVLTPE 355

DB 317 SFQTKRADSPVAGTATFTFECTLWGPSPAYFWKSGQDLPSTVSDAGTKVSP-- 374
 QY 356 GRSVLISARFADDESGKVYTCNALVAGVSRRTVV 391
 DB 375 --GTLTIEEVQVDEGATV-CAGMNSAGSSLSKAL 407

RESULT 14
 ID 001632 PRELIMINARY; PRT; 874 AA.
 AC 001632;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE CODED FOR BY C-ELEGANS CDNA CEESC12R.
 GN ZK377.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightening J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproot J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Nhan M., Hawkins J.;
 RL Submitted (FEB-1997) to the ENBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1997) to the ENBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the ENBL/GenBank/DBJ databases.
 DR ENBL: U98183; AAB52657.1; -.
 DR HSP: P56276; ITLX.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; Ig; 1.
 DR PRINTS: PR00014; FNTYPEII.
 SQ SEQUENCE 874 AA; 95861 MW; BC7270818D734C9 CRC64;

Query Match 9.0%; Score 658; DB 5; Length 874;
 Best Local Similarity 22.9%; Pred. No. 7.6e-38;
 Matches 229; Conservative 154; Mismatches 329; Indels 288; Gaps 35;

QY 400 PPPIEHQGVNQLPVKSVIPLPCRTLTGTPVQVSVLGDIPQVQHEHRLNSDAGALT 455
 DB 28 PPPIEHQGVNQLPVKSVIPLPCRTLTGTPVQVSVLGDIPQVQHEHRLNSDAGALT 455
 QY 460 ISDLQHEDEGLYTCVANSRNGKSSWSGVLRLDTIPNPNKIFRAPELSTPPTGGPKQM 519
 DB 519 ISDLQHEDEGLYTCVANSRNGKSSWSGVLRLDTIPNPNKIFRAPELSTPPTGGPKQM 519

```

RESULT 15
Q62845
ID Q62845 PRELIMINARY; PRT; 1026 AA.
AC Q62845;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE NEURAL CELL ADHESION PROTEIN BIG-2 PRECURSOR.
GN BIG-2.
OS Rattus norvegicus (Rat).

```

Query Match : 8.3%; Score 603; DB 11; Length 1026;
Best Local Similarity 23.4%; Pred. No. 7.5e-34;
Matches 275; Conservative 164; Mismatches 386; Indels 348; Gaps 45;

```

QY      2 ENPHRIEHPMOTTPVKNDPPTFNQABGNPTPIQWQDGRKEKLTUV--SHRILPAGGL 50
Db      30 QPESHVMPDSEKX-----VKLSCVGNKPKPHIKWLKNTDV--DIGDMFRYSVGGSL 84

QY      61 PFLVKLTHSRSDGADGVCCKANRFGVARSNAITQLQVAILRDEFLREPANT--RVAQGEVA 119
Db      85 L--INNPTQIQDSSTQCIATNSFCGISREAKIQFAL--ENFKTRTRSTVSVRGGQM 140

QY      120 LMBGAPAPGSPPEQIISWRNKGQTLNLVWNRIRITVDGNGALQEARQSDGQYGVVVKV 179
Db      141 VLICGPPPHSGSLSIAMIN--EHPYSQDNRRFVSQETGNLIYAKVEKADGWNTCVWNT 199

QY      180 VGTRESATAFLKVVWRPFLLRG-----PQNQTAVGSSVWQCRIGGSP 224
Db      200 VTSQ-----VLGPPTPLLRINDGMYETEPKIEVQFPETVAEKGSTVKLSCFALGNV 254

QY      225 PDVLNRRRTASGNMPLKFSMLBSASGRVHVLEDLSKLDVTDLMWGECTCEADNAVGG 284
Db      255 PTLNMR--ADG--KPIAKRRHKSQ-----LLEIPNQQDAGSTCEVANSRKG 303

QY      285 ITATGILVTHAPKPIFRPNKQLVEIGDEVLFECQANGRPPTLYKSWEGNSSLGPTLR 344
Db      304 NIAGQVFTFYAQPNWQIINDIHVAMEEYWFECNAGRPRKPIYRWLKGND-----PLLT 358

QY      345 DGRMEVILTPDGRSVLSIARFARSDSKVCTNALNAGVSSVSTRVSVVDQFLPPPII 404
Db      359 RERIQE-----QGTNLITVLSDAG--MYQCAENKKGVIYASAGSLV-----IA 403

QY      405 EQGQVWQTL-----PVKSIVLPCRTGLTPVPQVSWLDGIPDQVHEHRNLSDAGA 457
Db      404 EPDFSRSTLLKRVTLVKGVEVLECKPASPRPVYTW--RGRGT--LRENERITISDGN 461

QY      458 LTIISDQHEDEGLTYCVASNRNKGSSGYLRLOTPTNPNI----- 499
Db      462 LRINIVTK--SDAGSTCIATNHGTAASSTGNVGVDPKVKVWPPSMDVTVGESILPQC 520

QY      500 ----- 499
Db      521 VTHDSHSLVITPTWFMGHILIDFDKGDHFRVGGQSDGMLMRNLQIKHAGKVCWQVT 580

QY      500 -----KFFRAPELSTGPPGKPMY-----EKGENSEVLSWTRSNKVGSSLVGTVIEM--F 550
Db      581 SVKLKSAADI--TVKPPGPPPAVITDEITDITDIAQISW--RPGDNHSEITMYVIAQRTFP 638

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:17:58 ; Search time 233.01 Seconds
(without alignments)
190.332 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYTLGPGYHTHTHTYINFD.....TAQRFSPRNGIVTQQT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6860	100.0	1297	20 Y13565	C. elegans Robo po
2	6860	100.0	1297	20 Y08403	C. elegans ROBO pr
3	1588	23.1	1395	20 Y13563	Drosophila Robo 1
4	1588	23.1	1395	20 Y08401	Drosophila sp. ROB
5	1500.5	21.9	1651	20 Y13566	Human Robo 1 polyp
6	1489.5	21.7	1649	20 Y08404	Human ROBO1 protei
7	1350	19.7	1380	20 Y08402	Drosophila sp. ROB
8	1344.5	19.6	1381	20 Y13564	Drosophila Robo 2
9	1266	18.5	753	20 W83927	Human TRS protein.
10	600.5	8.8	1018	18 W06485	Rat contactin liga
11	600	8.7	1571	19 W42087	Human Down syndrom
12	598	8.7	1910	19 W42086	Human Down syndrom

13	593	8.5	1018	15 R63759	Human contactin (F
14	593	8.5	1018	17 R87028	Human contactin.
15	583.5	8.5	1257	20 W74152	Human L1 cell adhe
16	583	8.5	1028	19 W29667	Homo sapiens DCS
17	580.5	8.5	1225	19 W52289	Homo sapiens cdo t
18	561.5	8.2	434	20 Y13567	Human Robo 2 polyp
19	561.5	8.2	434	20 Y08405	Human partial ROBO
20	554	8.1	1447	16 R68553	Deleted in colorec
21	554	8.1	1447	20 Y33498	Human DCC protei
22	554	8.1	1728	12 R13144	Deleted in Colorec
23	546.5	8.0	1242	19 W52287	Rattus norvegicus
24	541.5	7.9	1192	19 W57900	Protein of clone C
25	539.5	7.9	1299	21 Y40439	Human Nr-CAN prote
26	526	7.7	1897	21 Y81785	Human protein tyro
27	526	7.7	1897	21 Y56100	LAR tyrosine phosph
28	524.5	7.5	1304	19 W59994	Human neural cell
29	518.5	7.6	1251	19 W37778	Rattus norvegicus
30	510.5	7.4	1496	20 W81030	Melanoma associate
31	510.5	7.4	1496	21 Y70469	Human p53 target n
32	508.5	7.4	4412	21 Y53666	Sequence gi/101742
33	501.5	7.3	3117	21 Y53667	Sequence gi/332818
34	495	7.2	1911	16 Y71726	Human PTP-DB. Hom
35	495	7.2	1911	18 W27225	Human protein tyro
36	495	7.2	1911	20 W94027	Human protein tyro
37	485.5	7.1	1139	19 W37779	Rattus norvegicus
38	485	7.1	1125	19 W52288	Rattus norvegicus
39	465	6.8	1501	16 R72858	Rat receptor type-
40	441.5	6.4	1070	18 W08747	Human colon carcin
41	439	6.4	761	17 R92255	Neural cell adhesi
42	434	6.3	1853	21 Y53668	Protein 608 sequen
43	434	6.3	2387	21 Y53665	Mechanical stress
44	434	6.3	2597	21 Y53664	Mechanical stress
45	432	6.3	848	21 Y88565	Human NCAM 140KD 1

ALIGNMENTS

RESULT 1
Y13565
ID Y13565 standard; Protein; 1297 AA.
XX
AC Y13565;
XX
UT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX
PW W09925833.1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.
 XX
 SQ Sequence 1297 AA;

Query Match 100.0%; Score 6860; DB 20; Length 1297;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYTLGFYHTHTHTHTYINFKIPNASLAPVIEHPDITVSRGSPATLACGAKPSTAKI 60
 DB 1 MYTLGFYHTHTHTHTYINFKIPNASLAPVIEHPDITVSRGSPATLACGAKPSTAKI 60
 61 TWYKQGPVITKEQVNSHRIVLDTGSLFLKWSNGKSDGADGAYCVASNEHGEVKS 120
 DB 61 TWYKQGPVITKEQVNSHRIVLDTGSLFLKWSNGKSDGADGAYCVASNEHGEVKS 120
 121 NEGSALALAKREDVRPRPTQALGGEMVLECSPPGPEPVVSWKDKELRIQDMR 180
 DB 121 NEGSALALAKREDVRPRPTQALGGEMVLECSPPGPEPVVSWKDKELRIQDMR 180
 181 TYLHSDGNLIIDVDRSDSGTYQCANNMVERGVNSPARLSVFEKPFKEPEKMDVDG 240
 DB 181 TYLHSDGNLIIDVDRSDSGTYQCANNMVERGVNSPARLSVFEKPFKEPEKMDVDG 240
 241 AAVLFDRCVTPGQPGQPIITWKKNEPMTVTRAYIAKDNRLRIERVQSDSGEYCTAARNP 300
 DB 241 AAVLFDRCVTPGQPGQPIITWKKNEPMTVTRAYIAKDNRLRIERVQSDSGEYCTAARNP 300
 301 AGTLESASLHQAAPPSPQTKPADQSPAGGTATFECTLVGQSPATYVSWKEGQDILLP 360
 DB 301 AGTLESASLHQAAPPSPQTKPADQSPAGGTATFECTLVGQSPATYVSWKEGQDILLP 360
 361 SYVSADGRVTSPTGLTIEVRQVDEGAYVCAQMSAGSSLSKAALKATFETKGVKQKK 420
 DB 361 SYVSADGRVTSPTGLTIEVRQVDEGAYVCAQMSAGSSLSKAALKATFETKGVKQKK 420
 421 KSKMGKQKQKQKQSIKYLISAFTGTPAKPPPTIEBHGQNTLMVSSSAILPQASGKP 480
 DB 421 KSKMGKQKQKQKQSIKYLISAFTGTPAKPPPTIEBHGQNTLMVSSSAILPQASGKP 480
 481 TPGISNROGLPDIIDTDSRISQHSSTGSLHIAKPKPTGVYTICLAKNDEGSTWSASLT 540
 DB 481 TPGISNROGLPDIIDTDSRISQHSSTGSLHIAKPKPTGVYTICLAKNDEGSTWSASLT 540
 541 EDHTSNAQVIRMPDPSFSPPTQPIVNVDTIEVELHWNAPSTSGAGPIGTIYITSP 600
 DB 541 EDHTSNAQVIRMPDPSFSPPTQPIVNVDTIEVELHWNAPSTSGAGPIGTIYITSP 600
 601 DLQGTWFIIDPVASTEYRIKGLKPSHYMFVIRAEENKIGTGPSVSSALVTSKPAQV 660
 DB 601 DLQGTWFIIDPVASTEYRIKGLKPSHYMFVIRAEENKIGTGPSVSSALVTSKPAQV 660
 661 ALSDKNMKMAIEAKRLTSEQLIKLEEVKTINSTAVLFWKRLLEELIDGYIKWRGPP 720
 DB 661 ALSDKNMKMAIEAKRLTSEQLIKLEEVKTINSTAVLFWKRLLEELIDGYIKWRGPP 720
 721 RTDNDQVNVVTSPTSTENTVVSMLMFTNYEFFVPHSGVSHIGAPSNMDVLTAEP 780
 DB 721 RTDNDQVNVVTSPTSTENTVVSMLMFTNYEFFVPHSGVSHIGAPSNMDVLTAEP 780
 781 SLPPEDVIRMLNLTIRISWAKPADGINGLKGPIVIVQAPNNRNTTNEASV 840
 DB 781 SLPPEDVIRMLNLTIRISWAKPADGINGLKGPIVIVQAPNNRNTTNEASV 840

QY 841 TLFLHVTGMYTKIRVAARSNGGVSHGTSSEVIMNDQTLKHLAAQGNESFLYLINKS 900
 DB 841 TLFLHVTGMYTKIRVAARSNGGVSHGTSSEVIMNDQTLKHLAAQGNESFLYLINKS 900
 901 HVPVIVTALLIIFVVIILAYCTWNSRSDGKRSFINKDGSVHMASNLDVQAPN 960
 DB 901 HVPVIVTALLIIFVVIILAYCTWNSRSDGKRSFINKDGSVHMASNLDVQAPN 960
 961 QNMYTATGMYTMNNRNGALYSLPNAQDFPNCCDYSGTHRPGSEHYHAYLTGGP 1020
 DB 961 QNMYTATGMYTMNNRNGALYSLPNAQDFPNCCDYSGTHRPGSEHYHAYLTGGP 1020
 1021 GNAMSTFYGNGYHDDPSYATTTLVLSNQPAWLNKMLRAPMPTVPPPEARTADH 1080
 DB 1021 GNAMSTFYGNGYHDDPSYATTTLVLSNQPAWLNKMLRAPMPTVPPPEARTADH 1080
 1081 TAGRERSRSRSDGRTGLMGLHRTSGSQRSDSPPTDVTYVLHSSDGTGSSKERTGE 1140
 DB 1081 TAGRERSRSRSDGRTGLMGLHRTSGSQRSDSPPTDVTYVLHSSDGTGSSKERTGE 1140
 1141 RRPPTAKTLMFIPPPSPNPPPGGVHVDYATRLQNLNKGSTPREDTYDVSQGAFAVD 1200
 DB 1141 RRPPTAKTLMFIPPPSPNPPPGGVHVDYATRLQNLNKGSTPREDTYDVSQGAFAVD 1200
 1201 NARPTSNMNLGDLPLAGKRDGSDGSSLDGSDGSEADGEMSESDVPRGVKAVPR 1260
 DB 1201 NARPTSNMNLGDLPLAGKRDGSDGSSLDGSDGSEADGEMSESDVPRGVKAVPR 1260
 1261 MGISASTLAHSCYTGINSTAGTAFRSIPRNGVITVROT 1297
 DB 1261 MGISASTLAHSCYTGINSTAGTAFRSIPRNGVITVROT 1297

RESULT 2

Y08403
 ID Y08403 standard; Protein: 1297 AA.
 XX
 AC Y08403;
 XX
 DT 24-JUL-1999 (first entry)
 XX
 C. elegans ROBO protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 KW cell morphology; screening assay.
 XX
 OS Caenorhabditis elegans.
 XX
 PN W09920764-A1.
 XX
 PD 29-APR-1999.
 XX
 PP 20-OCT-1998; 98W-US22164.
 XX
 PR 14-NOV-1997; 97OS-0971172.
 PR 20-OCT-1997; 97OS-0062921.
 XX
 PA (RBC) UNIV CALIFORNIA.
 XX
 PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
 XX
 DR WP1; 1999-312615/26.
 DR N-PSDB; X57252.
 XX
 PT Robo polypeptides, a new immunoglobulin superfamily member
 XX
 PS Claim 1; Page 52-63; 80pp; English.
 XX
 CC This invention describes novel Robo (roundabout) polypeptides. Involved
 CC in nerve guidance which have been isolated from Drosophila sp.,
 CC C. elegans, human and murine samples. The products of the invention can
 CC be used to raise anti-Robo antibodies, which can be used to modulate cell
 CC function or morphology. The Robo polynucleotides and fragments are useful

CC as probes and primers and for production of the Robo polypeptides. The
 CC probes and primers are also useful in screening assays.

XX
 SQ Sequence 1297 aa;

Query Match 100.0%; Score 6860; DB 20; Length 1297;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYTLGFTHTHTHTHTYINFDKIPNASNLAPVIEHPDVLVSVRGSPATLACQAKSPSTAKI 60

Db 1 mytlgfythththtyinfdkipnasnlaviehpdlvsvrgspatlcqakspstaki 60

Qy 61 TWYDQGPVITNKQGVNSHRIVLDTGSLPLKLVNSGKNGKSDAGATCYVASNEHEGVKS 120

Db 61 twydgqpvitnkgvnsshrivldtgsllkvlvnsngkngksdagatcyvasnehegvks 120

Qy 121 NEGLSKLAMLREDFVRPRTVQALGSEMAVLECSPPGPEPVNSKDDKELRIQDMPR 180

Db 121 neglsklamlredfvrpvtqalgselavlecsppgpepvnskddekrlrqdmp 180

Qy 181 YTLHSQNLIDPVDSDSGTYGCVANMVGERSAPARLSVFEKPFEPKMDTVQV 240

Db 181 ytlhsqnlidpvdsgdsgtygcvanmvgersaparslvfekpfepkmdtvqv 240

Qy 241 AAVLFCRVTVGDPQPOLITWKRKNPEPVTRAYIAKQNRGLIERVQPSDEGEYVCTARNP 300

Db 241 aavlfcrvtvgdpqpoltwkrknpepvtrayiaknrgliervqpsdegeyvcyarnp 300

Qy 301 AGTLEASAHLRVQAPPSFQTKPADQSVAGGTATFECTLVQGPSPAYFWSKEGGQQLLFP 360

Db 301 agtleasahlrvqappsfqtkpadqsvaggtatflectlvqgpspayfwskeggqqlf 360

Qy 361 STYDAGDKRVTGSPGTGLTIEVRQVDEGAYCAWNSAGGSLKAAKATFETKRGVQK 420

Db 361 stydagdkrvtgsptgtltievrqvdegaycawnsaggslkaakatifetkrgvqk 420

Qy 421 KSMGKQK 480

Db 421 ksmgkqk 480

Qy 481 TPGISWLDGLPIDIDTSRISQHSSTGSHIADLKPDGTQVTCIAKNEDEGSEWASLTV 540

Db 481 tpgiswldglpididtsrisqhsstgshiadlkpdgtqvytciaknedegeewasl 540

Qy 541 EDHTSAQVNMVQMPDPSNPPSSPTQPIIVNVDTVELVHNAFSTSGAGPTGIIYITSP 600

Db 541 edhtsaqvnmvqmpdpsnppssptqpilvvnvdtvelvhnafstsgagptgiilytsp 600

Qy 601 DLQGTWFNPIDVASTETRYIKGLKPSHSYMFVIRNAEKIGTSPSALVTSKPAQV 660

Db 601 dlqgtwfnpidvastetryikglkpshsymfvirnakeigtspsalvtskpaqv 660

Qy 661 ALSDKKMDMAIAEKRLSEQLIKLEEVKTINSTAVRLWKKRLEELIDGTYIKWRGP 720

Db 661 alsdkkmdmaiaekrlseqlikeevktinstavrlwkkrlleelidgtyikwrgp 720

Qy 721 RINDQNVVNTSPSTENTYVSNMLPFTYEFFVPIHSGVHSIGAPSNMDVLRAEP 780

Db 721 rindqnvvtspstentyvsnmlpftyeffvpihsgvhsigapsnmdvlraep 780

Qy 781 SLPPEDVRIRMLNLTIRLSWKAPKADGINGILKGQIVIVQAPNNRNITNTERAASV 840

Db 781 slppedvrirmnlntirslwkapkadgingilkgqivivqapnnrnitnteraasv 840

Qy 841 TLFHUTGTYTKIRVAARSNGGVSHSTSEVIMNDQLEKLAAQOENESFLYGLINKS 900

Db 841 tlfhutgtytkirvaarsnggvshstsevimndqleklaaqoenesflylinks 900

Qy 901 HVPVIVIVAILIIFVWIIAYCTWRNSRSGKDRSFIKINDGVSVMASNLWDVAQNP 960

Db 901 hvpvivivailiifvwiiaaycwrnsrsgkdrsfikindgsvmasnlwdvaqnp 960

Qy 961 QNPMTNTAGMTMNNRNGALYSITPAQDFNWCDDYSGTMRGPSEHGHYAQLGGP 1020

Db 961 qnpyntntagmtmnnrngalysitpnadfnncddysgtmrhpseghghyaqlggp 1020

Qy 1021 GNASMTFTYGNQTHDDPSPTATTLVLVSQPAWLNKMLRAPMPTNPVPPPPARYADH 1080

Db 1021 gnamstfytgnythddpsptattvlvsqpwlnkmlrapmntnpvpppparyadh 1080

Qy 1081 TAGRRSSRSRSDGRTGLNGLHRTSGSQRSDSPHTDVSYVQLHSDGTSSEKERTGE 1140

Db 1081 tagrrssrsrdgrtgleghlhrtsqgsrdsphtdvsyqlhsdgtsskertge 1140

Qy 1141 RRTPPNKTLMDFIPPPSPNPPPGHVYDTATRRQLNRGSTRPREDYTSVSDGAFARVDV 1200

Db 1141 rrtppnktlmdfipppspnpppgvhdydtarrqlnrgstrpredytsvsgafarvdv 1200

Qy 1201 NARPTSRNRNLGGRPLKGRDSDSQRSLMDDDGSSSEADGENSEGDPVPGGVKAVPR 1260

Db 1201 narptsrnrnlggrplkgrddsdqsrslmdddgssseadgensegdpvpggvkavpr 1260

Qy 1261 MGISASTLAHSCYGTNGTAQRFRSIPRNNIVTQBOT 1297

Db 1261 mgisastlahscygtngtaqfrsiprnnivtqbot 1297

RESULT 3

Y13563

ID Y13563 standard; Protein; 1395 AA.

XX

AC Y13563;

XX

DT 30-JUL-1999 (first entry)

XX

DE Drosophila Robo 1 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;

KN modulation; nerve cell function.

XX

OS Drosophila sp.

XX

PN W09925833-AL.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98NW-US24327.

XX

PR 14-NOV-1997; 97OS-0065543.

XX

PA (RBGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55767.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX

PS Disclosure; Page 30-33; 56pp; English.

XX

CC The invention relates to a method for modulating the amount of Comm
 CC (commissureless) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo/Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.

XX

SQ Sequence 1395 AA;

1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 2680, 26

Query Match 23.1%; Score 1588; DB 20; Length 1395;
Best Local Similarity 31.0%; Pred. No. 2.1e-89;

Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41.

[illegible]

```

QY 1046 -----LSQQQA-KWLNCKRARPAPMT-----PVPEE-PPARYAD 1079
      : : : : : : : : : : : : : : : : : : : : : :
Db 1059 siadckdsgtshpsydafagqcpavpvksnylqypvewlnseflppphhppsstys 1118
      : : : : : : : : : : : : : : : : : : : : : :
QY 1080 HTAG-RKRSRSRADGGR-----TLNGLHHRTSSQRS-----DSPPHTDWSI 1122
      : : : : : : : : : : : : : : : : : : : : : :
Db 1119 yaggspestrkskagsgistngslhshlsssgsgfawsgqyavacpennysn 1178
      : : : : : : : : : : : : : : : : : : : : : :
QY 1123 VQLHSSDGTGSKSKERTGERITPNKTLMDIPPPPPNPPPGGHVYDTRAQ----- 1175
      : : : : : : : : : : : : : : : : : : : : : :
Db 1179 plsaavgvnyrqlptngppqlpay-fatggagavp-nhl-pfatqahaseyga 1235
      : : : : : : : : : : : : : : : : : : : : : :
QY 1176 -LNRGTFREDTYS-----VSDGAFARVDYNA-RPTSRNRNL----- 1200
      : : : : : : : : : : : : : : : : : : : : : :
Db 1236 glnaarcagsgsrncsdalatpsmqppppvpegvqghvqshpmhtssnhlyqc 1295
      : : : : : : : : : : : : : : : : : : : : : :
QY 1212 GGRPLRGKRRDDQSRSLSMDDDGSSSDAGNSESNDVP 1250
      : : : : : : : : : : : : : : : : : : : : : :
Db 1296 sscsdshrsrscskrolcleghssakorgchrrrao 1334
      : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 5
YI3566
ID YI3566 standard; Protein; 1651 AA.
XX
AC YI3566;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
PN WO9925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REBC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-338008/28.
DR N-PSDB: X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1651 AA;

```

Query Match 21.9%; Score 1500.5; DB 20; Length 1651;
Best Local Similarity 32.1%; Pred. No. 6.5e-84;
Matches 408: Conservative 166: Mismatches 480: Indels 219: Gaps 36:

Qy **30** PVIIEHPIDVVVSRGSPATLNCGAK-PSZKITWYKDDQPVITNKEGVNSRHLVDTSGL 88
 |::||| :||| | |||| : | || :: | :|| :

Db **68** privehpdsllvskgepatlnckaegrppttiewykkgervetdkdopshrmlpsgl 127

Qy **89** FLKVNSGKNKDSDAGAYIVCASNEHGEVKSNEGSLKLAMREDFVRPRPTQALGGEM 148

```
Qy 1155 P P P S N P P P ----- P G G H V Y --- D T A T R R Q L N R G S T P 1162
      |||::|||           | :|           : |||
Db 1186 p p a h p p h s n s e e v n i s v d e s y d g e m p c v p a m y l g a d e l e e e d e r o n t o 1239
```

Query Match 19.7%; Score 1350; DB 20; Length 1380;
Best Local Similarity 27.0%; Pred. No. 9.le-75;
Matches 386; Conservative 217; Mismatches 494; Indels 334; Gaps 45;

Qy	30	PLIIHEHLDVWSRSGPALNCGAKPS-TAKITWKKDQPVITNKQVKSIRLVDLTSGL	68
Db	4	prliehgmddtvpkndpfncnqagndpttqivkdgrel-----ktdgshrnlpaggl	60
Qy	89	FLLKVSGKWKDSDAGATYCVASNEHGKVSNEGSLKMLALBDFVRPRTVQALGEM	148
Db	61	ffkivhser--resdgtwceakngfgvarsnatqlgavrldeflepanrtvawg	118
Qy	149	AVLECSPPRGPEPPVWSKDKKDLRQIDMPRTYSLASGDLNLDVDRSDSGTQVAMN	208
Db	119	almeacagprgspeglsxknngntlnlygnkrifvdggnlaieqarugddgrcyovkn	178
Qy	209	MVGERSVPAHLSVPEKPKFEGDKMTVDGAAVLFCDRVTGDPQGPQITWKR--KNEFM	266
Db	179	vvgtrrestatfkivhwrflrllpgnqgtavysvsvfciqriddgldpdlvrrtassgm	238
Qy	267	PVT-----RATAKONGRLIRSEVPSDEGEVVCYARWPAQTALSAHLVQAPD	316

```

Db      239 plrkfwlshasgrvhtvelrskldlvleddemeytceadnavggitatgilrvhapp 297
QY    317 SPQTIPADSPSVAGGATATEPCTIQCPSRAYTMSKEGGDLLFSTPVSDARKTKVST-   374
       | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     298 kfvirpkqnlvelgderlfecagqhprplysvsgnsillpgy--rdgmewtlpte 355
QY    375 ---GLITIEEVQQVDGAIV-CAGMNSAGSSLSK-AALKAFPTGTGRVKKSCKMGKQG 429
       | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     356 grslvislarfaredgegwvtcnalvwgvsstvrsvwtgfel-----           399
QY    430 KMWQSIIKYILIASVTGNPAKFPPPTIEHGKNQLWVGSSAILPCQAASGKPTPGISWLRD 498
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     400 -----pppliegvgpnqlpvksitwlpclrtglptqvswyld 438
QY    490 GLPIDITD-SRIQSHTGSLSHAIADLK-PDTGVYTCIAKNEDGESTWSASLTVEHTS-N 546
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     439 glpidvghernalsdagaitlsidqhedegelytcvasnngksswgvyldrlotups 498
QY    547 AQRVPMPDPNSESPTPIIVNVTDREVELHNWNPASTSGAGTGTOTITIQYTSLDLQWT 606
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     499 ikffrapeisltpypggkpqmvegensvltsrwnkvgsilgvlyiemfgmktedgy 558
QY    607 FNIPDTPVASTETIRKLPHSHSYMFIRAENEGKITGPSVSAAVLTTSPAQAVALSKDN 666
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     559 vavgtrvcntqtigtqlpgvnyfliraenshgilspsmspetivtryfnsgl---- 614
QY    667 KMMDAAREKRLTSBQLIKLEVYTIINSTAVLFWKKRLEELIDGYTIKW----- 717
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     615 ----disearaslsgdvnelsnasvdtsmkltwqlin-gkyvegilyvarqlpinnv 671
QY    718 -GPPRTNDNQYNVNTFS-----                               734
Db     672 patsntcnpilgsttsasasasallstkpnlaasekrdrgetnsggggatpletcky 731
QY    735 -----TENVVSNLMNPITYEFPYPFHVSHTSIHAPSNSMDVLTAEPSSLPP 784
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     732 rmltilingggasscttgllqvgtlylefviptyfk---evegkpsnrirtledvseap 788
QY    785 EDVRIMNLMLTLIRISKWAPKDINGILKGFIIVTG-QAPHNNR---NITTNRAS 839
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     789 ymeallinsavfilkwapeklrdhrylnhyvirvgdidahnfriltvntdiastp 848
QY    840 VILEFLVTGMTYKIRVAARSNGGVVSHGSTSEVMQNDELKHLLAQOENESFYGLINK 899
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     849 lylanitegvmytvgvaannagvpy--cvpatrlridtkrl-----dfpin-q 897
QY    900 SHVP-----VIVIVAILLVFVIIAYCYWNRSNDSGDGRSPIKINDGVSH-WASN 950
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     898 dhndvdtcpwflllgailgalvnlisganvfyk-----rhmmmqkelintargn 948
QY    951 NLWDVAQNPNPNQMNYTAGRMNMNRNMQALYLTPNAQDFFNCDIYSZTHMRG---- 1006
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     949 htsdvdkmps-----lsarngyvldsst-----gwmwrpspggd 985
QY    1007 ---SEBHRYLAQLGPGN-----                                1022
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     986 slenckhiadiaypcvagapspeggtssaggsgagsgdgidhhbgssernqqryyg 1045
QY    1023 -----AMST-----YGNOQDDSPSTATTLVLNSQCPAWLANDKLRAPMPT 1066
       ||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1046 eysniptdyevsfsgkapesyrgmgaspapyataslsiphaq-----qqqqqrpyqg 1099
QY    1067 NPVP-----PEPPARTADHTAGRRRSRNASDGRTLGNGLIHERTSQSQRSDPHHYD 1120
       ||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1100 rprryyglylrpmhp-----hyqqqhqhqhqgaq-----thqhqbalqhqhlppsnl 1146
QY    1121 SYVLQHSSD-----GTGSSKERTEGR-KTPNNPLMDPFPPPPSNPPPGEVITVATR 1174
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1147 yqgmtstseyiptngsrserwysegyyppdkzrhikhenkln-----chtyeaapr 1200
QY    1175 QLNQRSCTPRETDYSDSGAFARVDVNAPTSENRLNKGLRGPKLKG-RDDDSQRSLAMD 1232

```

Dn 1201 ---kgsspiisqfas-----vrgqlppncsigresarfkvintdggknoqnild 1247
Qy 1233 DDGGSSEADGENSEGDPVPGGVKXAVPRMGISASTLA-----HSCYGT 1275
Db 1248 ldsasmcyncladsq-----causspmamlmshehdahlyht 1285

||| : ||
|| | :

```

RESULT      8
Y13564
ID          Y13564 standard; Protein; 1381 AA.
XX
XX
AC          Y13564;
XX
XX
DT          30-JUL-1999 (first entry)
XX
XX
DE          Drosophila Robo 2 polypeptide.
XX
XX
KW          Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW          modulation; nerve cell function.
XX
XX
OS          Drosophila sp.
XX
XX
PN          WO9925833-A1.
XX
XX
PD          27-MAY-1999.
XX
XX
PF          13-NOV-1998; 98WO-US24327.
XX
XX
PR          14-NOV-1997; 97US-0065543.
XX
XX
PA          (REBC ) UNIV CALIFORNIA.
XX
XX
PI          Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
XX
DR          NPI; 1999-338008/28.
XX
XX
DR          N-PSDB; X55768.
XX
XX
PT          Modulation of Robo-Comm polypeptide interactions
XX
XX
PS          Disclosure; Page 34-38; 56pp; English.
XX
XX
CC          The invention relates to a method for modulating the amount of Comm
CC          (commissureless) polypeptide in contact with a cell expressing active
CC          Robo (roundabout) on its surface. The method comprises modulating the
CC          effective amount of Comm polypeptide in contact with the cell, where the
CC          amount of expressed active Robo is specifically modulated inversely with
CC          the modulation of the effective amount of Comm in contact with the cell.
CC          The method is used to modulate the amount of active Robo expressed on a
CC          cell. The method can be used to screen for agents that modulate Robo:Comm
CC          interactions. This is particularly useful for modulating nerve cell
CC          function.
XX
XX
SQ          Sequence 1381 AA.

```

Query Match 19.6%; Score 1344.5; DB 20; Length 1381;
Best Local Similarity 26.8%; Pred. No. 2e-74;
Matches 386; Conservative 216; Mismatches 486; Indels 353; Gaps 45

QY	30	PVILIEHPIDVSVSRGSPATLKGAKPS-TAKITWYKDGQVINKQWRSIRVLVDGSL	88
Db	4	prliehpmdtvlvknpgdfnfncqsegnptqlqfkgdrel---ktdtgsrhlmpaggl	50
QY	89	FLKLVSKGKNGKSDAGATYCVASNEHGKVSNEGSLKMLMLREDFVRPPTVALGEM	148
Db	61	ffkvlshsr--resdsgtycpeaknfgvrsrnatlqgvldrefrlpeantrvagve	118
QY	149	AVLECSGPRGPEPVFVSVSRKDDKELRQIDMPRTYLSHSDGLLIDVDSGSGTQVANN	208
Db	119	almeceprgpeppelswknqgtlilvngkrirfildvgnlaigearcsddgrycvrwn	178
QY	209	MVGKVSIPALRSVFEKFKFDEPKMTVDSGAALVDCRVTGQDPQPIQWR--KNEPM	266

PT

```

RESULT 9
W83927
ID W83927 standard; Protein; 753 AA.
XX
AC W83927;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human T85 protein.
XX
FE T85; FHEB-604; FHEV-SD4; human; neurological disorder; therapy;
RW diagnosis.
XX
OS Homo sapiens.
XX
FH Key :Location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Protein 21..753
FT /label= Mat_protein
FT Region 525..610
FT /note= "has homology to a fibronectin type III
FT domain"
FT Region 638..727
FT /note= "has homology to a fibronectin type III
FT domain"
FT Region 43..101
FT /note= "has homology to a Ig superfamily domain"
FT Region 143..203
FT /note= "has homology to a Ig superfamily domain"
FT Region 237..298
FT /note= "has homology to a Ig superfamily domain"
FT Region 329..394
FT /note= "has homology to a Ig superfamily domain"
FT Region 433..491
FT /note= "has homology to a Ig superfamily domain"
FT Peptide 247..249
FT /note= "RGD motif"
FT Domain 516..600
FT /note= "cytokine receptor homology N-terminal
FT domain"
XX
PN W09848051-A2.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07714.
XX
PR 10-OCT-1997; 97US-0062017.
PR 18-APR-1997; 97US-0044746.
XX
PA (MILL-) MILLENNIUM BIOTHERAPUTICS INC.
XX
PI Holtzman D, McCarthy SA;
XX
DR WPI: 1999-024021/02.
DR N-PSDB: V59278.
XX
PT New isolated human FHEMA-070 and T85 proteins - used to develop

```

Matches 255; Conservative 124; Mismatches 374; Indels 323; Gaps 39;

Qy 4 LGFYTH-THTHTINFDKIPNASLAPVIEHPIDVVSRS---PATLNQAKPSTAK 59
 Db 19 Lefthvhrjyghvseedk----gglfifecpntlypeeslegkvslnccraspfp 73
 Qy 60 LTKY----DGPVITNKKEQVNSHRIVLDTGSLFLKLVNSGKNGKSDMAGATYCVASNEH 115
 Db 74 --vkwmnqgdvlta-----drysmggnlvi-----nnpdkqdaqlyyccasnny 120
 Qy 116 GEYKSNESGLKAML----REDFVRVRPTQALGEMAVLECSPPRGFPPEPV-VSWKDD 170
 Db 121 gmvrtatlsfgyldpfpdp---rpe-kvkkgkmgvildcpypfpdpdlsyrlwef 176
 Qy 171 KEALRIQMPRTYLTSDGNLIDPVDSDSGTYCVANNMVGERSNPA-RLSVFEK---- 225
 Db 177 pfvitatdkrfvstqgnlyanvessdrgnyscf-----vsspsltkvsfkipl 228
 Qy 226 ---PKFQEP-----KMTVDVGAALVLCFRCVTPDPOPIITWKRKKMEPMVTRAYI 273
 Db 229 lpiptertkypadivvgfkdiymmggnvlecfalgpnvdrkvwkvepmptt-ael 287
 Qy 274 AKNDRGLIEHVQSDSGEYKCYARNPAGTLESAHLRVQAPSPQTKPADQSVAGGTA 333
 Db 288 stsgvakiifnlgledglyceaanirgkdkharyvqafpewehindetvdlgsdl 347
 Qy 334 PFETCIQMPSPATYFNSKEGQDLPLPSVYSDGRTKVSPTGTLTIEVYVQVDEGATYCA 393
 Db 348 ypwcvatgktpitrlwnky-----ayhkgelrlydvtfenagmyqci 391
 Qy 394 QMSAGSSLSKAALK-----ATPETKGRVQKKSKMGKCKKQKVSIKTLISAVGTWP 448
 Db 392 seaystyanaelkilaaptfe-----mngmkk-----lla----- 426
 Qy 449 AKPPPTIEBHQVQTLVMSGSSAILPQASGKPPGISVLRDLGLDITDSRISQSGTSL 508
 Db 427 -----kgvrilackpaagpafgskv-gtevlvsnrllwedgsl 468
 Qy 509 HIALAKPTQGVYTICIAKNEGSGTWSASLVDEHT-----SNAQVFM 552
 Db 469 einmttndgdytfaennrgkanstgltvntprillapinadtvygenatqcas 528
 Qy 553 PDPS-----NF----- 558
 Db 529 dplsditdwsfnvgvdfnkethihygrnfhidangellimnaglkhagrytctagt 588
 Qy 559 -----PSSPQPIIYNVDTIEVHLMNASTGSGATGTYIQTYSPL 602
 Db 589 lvdnsasadiivrgpppggplriedratsvltwsgdshs-plskytig-ktkl 646
 Qy 603 QGTWNTI---POYVASTERYNG--LKPSSHYMFIARNEKIGITPVSSVALTYSKPA 657
 Db 647 sddvkdaktdppliegmesakavdlipmeyerfvattltgtgepispnrlkdga 706
 Qy 658 AQVALSD-----KNKMDATAEKLRLSBLKLE----- 686
 Db 707 povapsdvvggggnreltltwapsreyhngnfyvafkpfdgevskkvtvntpdgt 766
 Qy 687 -----EWTINSTAVR 697
 Db 767 ryyhktetpstaftgvkvafnkgdgpysliavinsaqdapseaptvgyvlssels 826
 Qy 698 LFWKKRKLLEDGTYIWKRGPPRTNDNQVNVNTPSTENTV--VSNMPTTFTEFFYP 755
 Db 827 vhw-khvlekivesyqiraghdkeaaahrvqvs---qeysarlenllpdtqfyfiewga 882
 Qy 756 YHSYGSHIGASPNSMDVLRAPSLPPEDEVIRRMNLTLR-----ISWAPKADGI 809
 Db 883 cns---agcpgssdvietfrkappsgp---rl---lsavrsgrsyfytvhwalsn 932
 Qy 810 MGILAGPQIVVIGQAPNNNRITINERASVTLPHVTGMYIKYIARASNGGCV 865
 Db 933 estvtgkyilrpdgqgdklfstkhhsiev---lprdgeyvewrshdgdyg 985

RESULT 11
 W42087
 ID W42087 standard; Protein; 1571 AA.
 XX
 AC W42087;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Human Down syndrome-cell adhesion molecule DS-CAM2.
 XX
 KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; human.
 XX
 OS Homo sapiens.
 XX
 PN W09817795-AL.
 XX
 PD 30-APR-1998.
 XX
 PF 23-OCT-1997; 97WO-US19547.
 XX
 PR 25-OCT-1996; 96US-0029322.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Koberberg JR;
 XX
 DR WPI; 1998-271791/24.
 DR N-PDB; V31988.
 XX
 PT New Isolated Down's Syndrome-cell adhesion molecule - used to
 develop products for detection, diagnosis and therapy of
 developmental and neurological abnormalities
 XX
 PS Claim 2; Page 90-95; 109pp; English.
 XX
 CC This polypeptide comprises Down syndrome-cell adhesion molecule
 CC DS-CAM2, an extracellular soluble protein belonging to a novel
 CC subclass of the Ig superfamily with highest homology to neural cell
 CC adhesion molecules. Its amino acid sequence was deduced from cDNA
 CC clones (see V31982) isolated from a trisomy 21 foetal brain library.
 CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and
 CC lacks the entire transmembrane domain of DS-CAM1. The invention
 CC provides human and murine DS-CAM nucleic acid sequences (see also
 CC V31981, V31985-87), expression vectors and host cells, transgenic
 CC animals, antibodies, antisense oligonucleotides, and primers
 CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
 CC with developmental and neurological processes. They can be used in
 CC e.g. neural prosthetic devices used in entubulation methods of
 CC repairing (regenerating) damaged or severed peripheral nerves, and
 CC also in bioassays to identify agonists and antagonists. The products
 CC can also be used in detection, diagnosis and therapy of developmental
 CC and neurological abnormalities such as Down syndrome, mental
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,
 CC or schizencephaly.
 XX
 SQ Sequence 1571 AA;
 Query Match 8.7%; Score 600; DB 19; Length 1571;
 Best Local Similarity 25.1%; Pred. No. 1,4e-28;
 Matches 228; Conservative 129; Mismatches 373; Indels 180; Gaps 38;
 Qy 16 YINFDKIPNASLAPVIEHP-----DVVVSRSPTALNCGAKPS-TAKITWYKDD 66
 Db 386 fvrkkl-saagyvqvrvldgtpklsifasekvspaeypslmncvktgtpitltldd 444
 Qy 67 QPVITNKKEQVNSHRIVLDTGSLFLKLVNSGKNGKSDMAGATYCVASNEHGEYKSNESGLK 126
 Db 445 dplksgshrsqmitsegvlysniss---sqvrdgyvrrctannasgvr----- 493

Qy 127 LAMLEDFR---VWP-RVQALGEMAVLECSPPRGFPFVSWNRDDEKELRIQDMRY 181
 Db 494 lyqarlvwvgsasirpknkitalagrdrtylncr-vlgyppysikwyknsallpfahr-qv 551

Qy 182 TLASDGNMLIDPDVRS-DSGTYQCVANNWGERVSN-----PARLSVTEKPKBQ 230
 Db 552 afennqtklksdvqkvedgeytc--nvlvpqsltsqsvhvtkvppflqpfafprf-- 607

Qy 231 EPKMTVDVGAALFDC-RVTGDPDQPIWKKRNEPMPVTRAYIAKN-----RGLRIERV 285
 Db 608 -----sigqrvfipcrvvsqdlpittitwkdgrppgsly-wtldnidftslrisnl 659

Qy 286 QPSDEGEYCYVARNPAGTLEASAHLRVQAPSPQTKPADQSVPAAGTATFECILWGQSP 345
 Db 660 slmhngpytciarneaavehsqslivrvppkfvpqrdqgilygkavilncseegypv 719

Qy 346 AYPW--SKRGQQLLPSPVVSADGRKVSPTGLTIEEVROVDGATVCAGMNSAGSLS 403
 Db 720 tlwfksgagvqfqp--lalgrlqvlansllkhvveedsgylickvsnvqadv 777

Qy 404 KAALKATPTEKRWQKKSKGKQKQKQVQSIKLYLSAVTGNTPAKPPPTIEHGQNT 463
 Db 778 ks-----myltvki-----pamitsypatt 797

Qy 464 LMW-GSSAILPQASGKPTPGISLWRLDGLPID-----ITDSRISQHSOTSLHIAJLKK 515
 Db 798 latgggkmsctahgeklpivrvkedrilpenarylvstkegeevlslqilptvr 857

Qy 516 PDVGVTCIAKNEGEGSTWSASLIVEDHTSNAQVFMPPDSNFPSSQPIVNVNDETV 575
 Db 858 edsgfischainsfyedrgiilqtve-----ppdpeleikdvkart 901

Qy 576 ELHWNAPSTSGAGPITGIYIYQSPDLQGTWFI-----PDVASTETRIKGLAPSHS 628
 Db 902 lrlwa-gfgnspitlydle--cknksdwsdsgaktdvsglnsat--lidlhpst 955

Qy 629 YMFVIRAEKNGICTPSVSALVTISKPAQVALSKNMMAIKRLTSEQLIKLEEV 688
 Db 956 yslmyakr--lkspsenlitaadaepdg-----ppqevhle-- 994

Qy 689 KYTINSAVLKFW--KRLKE-LIDGYIKWGPPTNDNQWTV----TSPSTENTVVS 741
 Db 995 plssgsirvtwkpqkhklnqgiiryzqigr--eystgnqfnisvdsdgsdsevytlid 1052

Qy 742 NLMPPTNYEFVITHSVGHSIHGAPSNMDVITA--EAPPSLPEDVIRMANLTLTRI 799
 Db 1053 lnlkfqtgylvqpcura----gtgspqellittledvpeypenqalatspesisi 1107

Qy 800 SWKAPKADGINGILGQPIVIGQAPNNN-----NNITTHERAASVTLFHLVTGNTYKIRV 855
 Db 1108 swtstkealngilqfrviywanlmdgelgelknlttq--pslclgkleytaysyq 1165

Qy 856 AARSNGVGV 865
 Db 1166 lafrtagdv 1175

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..23
 FT /label= Sig_peptide
 FT Protein 24..1910
 FT /label= Mat_protein
 FT Domain 24..887
 FT /label= IG
 FT /note= "immunoglobulin type-C2 domain"
 FT 888..1594
 FT /label= FbN
 FT /note= "fibronectin type III domain"
 FT 1595..1616
 FT /label= Transmembrane
 FT 1617..1910
 FT /label= Cytoplasmic
 FT Region 24..126
 FT /label= Ig1
 FT Region 127..225
 FT /label= Ig2
 FT Region 226..316
 FT /label= Ig3
 FT Region 317..409
 FT /label= Ig4
 FT Region 410..506
 FT /label= Ig5
 FT Region 507..603
 FT /label= Ig6
 FT Region 604..697
 FT /label= Ig7
 FT Region 698..792
 FT /label= Ig8
 FT Region 793..887
 FT /label= Ig9
 FT Disulfide-bond 46..102
 FT Disulfide-bond 145..197
 FT Disulfide-bond 246..293
 FT Disulfide-bond 335..385
 FT Disulfide-bond 428..484
 FT Disulfide-bond 525..575
 FT Disulfide-bond 617..669
 FT Disulfide-bond 711..766
 FT Disulfide-bond 809..865
 FT Disulfide-bond 1307..1359
 FT Modified-site 78..80
 FT /note= "Asn is N-glycosylated"
 FT 106..108
 FT /note= "Asn is N-glycosylated"
 FT 470..472
 FT /note= "Asn is N-glycosylated"
 FT 487..489
 FT /note= "Asn is N-glycosylated"
 FT 558..660
 FT /note= "Asn is N-glycosylated"
 FT 666..668
 FT /note= "Asn is N-glycosylated"
 FT 710..712
 FT /note= "Asn is N-glycosylated"
 FT 748..750
 FT /note= "Asn is N-glycosylated"
 FT 795..797
 FT /note= "Asn is N-glycosylated"
 FT 924..926
 FT /note= "Asn is N-glycosylated"
 FT 1142..1144
 FT /note= "Asn is N-glycosylated"
 FT 1160..1162
 FT /note= "Asn is N-glycosylated"
 FT 1250..1252
 FT /note= "Asn is N-glycosylated"
 FT 1271..1273
 FT /note= "Asn is N-glycosylated"

RESULT 12

W42086

ID W42086 standard; Protein; 1910 AA.

XX

AC W42086;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM1.

XX

KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;

KW signal transduction; trisomy 21; mental retardation;

KW holoprosencephaly; corpus callosum agenesis;

KW schizencephaly; diagnosis; assay; human.

XX


```

RESULT 13
R63759
ID R63759 standard; Protein; 1018 AA.
XX
AC R63759;
XX
DT 10-MAY-1995 (first entry)
XX
DE Human contactin [EMBL Accession #221468].
XX
KW Human contactin; human brain glycoprotein; neural cell adhesion;
KW mouse F3, chicken contactin/Fil adhesion molecules.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 45..94
FT Disulfide-bond: 138..191
FT Disulfide-bond 243..290
FT Disulfide-bond 332..371
FT Disulfide-bond 416..464
FT Disulfide-bond 506..563
FT Domain 604..657
FT /label= FLR
FT /note= "Conserved core of fibronectin type III
FT like repeat"

```


Query Match 8.6%: Score 593: DB 17: Length 1018:

RESULT 15
W74152

ID W74152 standard; Protein; 1257 AA.
XX
AC W74152;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human L1 cell adhesion molecule.
XX
KW Human L1 cell adhesion molecule; LICAM; neurite growth;
KW nervous system development; nerve regeneration;
KW neuronal cell cohesive interaction.
XX
OS Homo sapiens.
XX
PN US5872225-A.
XX
PD 16-FEB-1999.
XX
PF 18-NOV-1994; J40US-0341843.
XX
PR 26-JUN-1992; J20US-0904991.
PR 18-NOV-1994; J40US-0341843.
XX
(UTCA-) UNIV CASE WESTERN RESERVE.
XX
Lennon V;
XX
PI
DR WPI; 1999-166719/14.
DR N-PSDB; X01598.
XX
PT Human L1 cell adhesion molecule - supports neurite outgrowth and is
PT involved in nervous system development and repair
XX
PS Claim 1; Fig 3; 45pp; English.
XX
CC This sequence is the human L1 cell adhesion molecule (LICAM) of the
CC invention. LICAM supports growth of neurites in vitro and is involved in
CC development of the human nervous system and in nerve regeneration. It is
CC useful in vivo and in vitro experiments on nerve growth and
CC regeneration. LICAM mediates cohesive interactions of neuronal cells to
CC each other and to extracellular matrix.
XX
SQ Sequence 1257 AA;

Query Match 8.54; Score 583.5; DB 20; Length 1257;
Best Local Similarity 22.94; Pred. No. 1.le-27;
Matches 243; Conservative 141; Mismatches 357; Indels 321; Gaps 42;

QY 30 PVIIEH-PIDVVYSRGSPATLCGA-KPSTAKITWKDG---QPVTNKEQV-----N 77
| | | | : | | | : | | | : | | | : |
Db 35 pviitqsprrlrvfpdtdislkcaesgkpe-vqfvrwdrghfkp---keelgtvtygs 89
| | | | : | | | : | | | : | | | : |
QY 78 SHRVLDTGSLFLKLVNSGNKGSDAGAYCVASNEHGVEVSNEGSLLKLMLREDFFVR 137
| | | | : | | | : | | | : | | | : |
Db 90 ph---sgsfittgnss--nfagrfgiyrcfasnklgtamshe---irlmaegapkw 138
| | | | : | | | : | | | : | | | : |
QY 138 P---RTYVALGGEMAVLCSPPRGFPPEPVSWRKDDKKELRIQMNPRTLHSGDNLIDP 193
| | | | : | | | : | | | : | | | : |
Db 139 pketkvprveeegsvlpncpnps--aeplriymnsksllihkgdvrtvmngqnlyfan 197
| | | | : | | | : | | | : | | | : |
QY 194 VDPSDS-GTYCVANNMVGKRVSNPARLSVFEXPKFQEPEKMITYDV----- 239
| | | | : | | | : | | | : | | | : |
Db 198 vltsdnhsdyichah-----fpgtrtliqk---epidrlvkatsmidrkpcrllf 244
| | | | : | | | : | | | : | | | : |
QY 240 -----GAALVFCGRVDPQPQIOTVKRRKNEMPMVTRAYLKADNRGLRIERVOP 287
| | | | : | | | : | | | : | | | : |
Db 245 otssashhlvcceupvlclclaeafotntklvrsosowcadrtvtvanhktilcllkvee 304

```

QY 428 SDGEVTEYRNAPGLEASALASLQVAPSFQTKPADQSVPACTATFCTCLWQQSPAY 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 edd9e9rciaenslgarhwyvtveasapwllkqshlysgpetarlcdqgypzrper 364

QY 348 FWSKQDQQLLFSTVSADGRVTVSPGCLILEEVQVDEGAYCAQMSAGSLSKAAL 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 twrlng-----lpwealcqdkyrlg-rallslsnvqsdmtvctcearrhlllanayl 419

QY 408 KATPTEKGRVQKKSQKMGKQKQKWSIIKLILSANTGNTGAPPPPIEHCNQWTLNW- 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 -----yvv-----glqkAlta-----dngtymav 439

QY 467 -GSSALLPQASQKSPGIGWL-ROGLPIDIDTISRISQHSSTGSLHIAIDLKPDGVVTCI 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 gqstalylckgfpappsvqwgldedgtv-lqderffpyangllsrlrdqandgryfcl 498

QY 525 AKNEDGESTWSALSTVEDBTSNACFVRMP-----DPSNFPs----- 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 499 aandqnmvtmanlkvkdattigtqgrstiekkgsrvftvcasfdpslqpsitvrgdr 558

QY 561 -----SPTQPIIWN----- 569

Db 559 digelgdsdkfyiedgrlvihsldysdgngyscvasteldesraqllrvsgspvprl 618

QY 570 -----VDTSEVELHWNAPSTSGAGPIGYIILQYSPDLG-QTWN--TPDVASTET 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 vlsldhlhtqgvrsv-spaechnaplekyldefedeknapexkyslqkvgpogattl 677

QY 619 RIKGLKPSHSTMFYIRNAENKIGTSPVSSALLTTSKPAQVAQSSOKNMQ----- 669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 k--lspvhytfrvtalnkysspgespsvstvtpe-----apeknprdvkgegnet 729

QY 670 -MAIAKRLTL-----SBOIL----- 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 nmritvklprlmdwnapqgyrvqrvpggrtgpweqelvsdplfvnsstfvrpeikvg 789

QY 684 -----KLEEVKINTSTAVRLPWKKRLLEL-----IDGYI 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 avnsqgkpgpvtqgtygysedypgaileglellnsavvlkwpvrdlqaykghlrgy 849

QY 714 IK-NR-GPPTNDNQTVN-----VTPSPSTENYSVNLMPPTNEFFVPIISGVHSIHGAP 767
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 850 vttyresqgrkshkrlhkhvvpantvslslsgrlyps-----yhlegafqfgrg 902

QY 768 SNSMDVLTAEPAPSLP--PEDVRIRMLNLTIRLSWAKPADGINSGLKGPVQIQVAP 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 903 sgspaefftsltpsgyqgpeahllcecgstslilwqplish-nvlytygyvisyphlde 960

QY 826 NNNRMITTEENAAVS-----TLFHLVTGMYTIRVAARSNGSGV 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 gqkqgslfnldrpehlrthaldtshprrfvgelattkeapd 1002

```

Search completed: January 22, 2001, 12:18:45
Job time: 1682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:25:18 ; Search time 325.28 Seconds
(without alignments)
270.743 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYTLGPGYTHHTHTHTYINFD.....TAQFRFSIPNNNGIVTQDQT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Indexed: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pirl:*
2: pirl2:*
3: pirl3:*
4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	6523.5	95.1	1273	2	742405
2	4628	67.5	874	2	T29548
3	2232	32.5	423	2	T29549
4	1505.5	21.9	1612	2	T30805
5	1483.5	21.6	1651	2	T14160
6	1323	19.3	1344	2	T14316
7	635	9.3	1028	2	I58164
8	632.5	9.2	1443	2	I50600
9	616	9.0	1028	2	A53449
10	607	8.8	1375	2	T13822
11	598	8.7	1896	2	T08851
12	597	8.7	1010	2	JU0094
13	595	8.7	1020	2	S05944
14	595	8.7	1021	2	A57112
15	593.5	8.7	1091	2	S01998
16	593	8.6	1018	2	A54744
17	592	8.6	1018	2	JC4211
18	589	8.6	1272	2	S26180
19	588	8.6	2222	2	T13924
20	585	8.5	1040	2	A34695
21	583.5	8.5	1257	1	A41060
22	582.5	8.5	1040	2	A49356
23	580.5	8.5	1240	2	T03097
24	580.5	8.5	1427	2	I51669
25	574	8.4	1259	2	A43425
26	567.5	8.3	1260	1	S05479
27	562.5	8.2	1036	2	S22383
28	561	8.2	1368	1	A39640
29	557	8.1	1259	2	S36126

30	554	8.1	1447	2	A54100	tumor suppressor p
31	553	8.1	1277	2	T30532	neural cell adhesi
32	552.5	8.1	1526	2	T13823	frazzled gene prot
33	550	8.0	1239	1	A32579	neuroglian - fruit
34	547	8.0	1256	2	T03096	CDO protein - rat
35	531.5	7.7	1197	2	T30581	neural cell adhesi
36	527	7.7	1907	2	S50893	protein-tyrosine-p
37	526	7.7	1897	1	TDHDLK	leukocyte antigen-
38	518.5	7.6	1898	2	S46216	leukocyte antigen-
39	509.5	7.4	1912	2	A56178	protein-tyrosine-p
40	508.5	7.4	1962	2	T38346	elastic titin - hu
41	501.5	7.3	5175	2	T20992	hypothetical prote
42	501.5	7.3	5198	2	T43290	hemocytin precurs
43	501	7.3	1894	2	C54689	protein-tyrosine-p
44	498.5	7.3	1232	2	T43027	neural cell adhesi
45	492	7.2	1863	2	S46217	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T42405
sax-3 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42405
R:Ballen, J.A.; Yi, A.A.; Bargmann, C.I.
Cell 92, 217-227, 1998
A:Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple
A:Reference number: 122160; MID:98117250
A:Accession: T42405
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <L>L
A:Cross-references: EMBL:AF041053; NID:g2804779; PIDN:AAC38848.1; PID:g2804780
C:Genetics:
A:Note: sax-3
C:Function:
A:Description: sax-3 function is required at the time of axon guidance

Query Match		95.14;	Score 6523.5;	DB 2;	Length 1273;	
Best Local Similarity		97.34;	Pred. No. 0;			
Matches 1243;		Conservative 1;	Mismatches 1;	Indels 33;	Gaps 3;	
Qy	24	NASNLAPVILIEHPIDVVS	RGSPATLMCAKPSAKITWYKDGQ	PVTINKEQVNSHRLV	83	
Db	25	NASNLAPVILIEHPIDVVS	RGSPATLMCAKPSAKITWYKDGQ	PVTINKEQVNSHRLV	84	
Qy	84	DTGSLFLKLVNSGKMGKSD	GAGAYCVASNEHGEVKSNEGSLKMLA	REDFRVRPTQVA	143	
Db	85	DTGSLFLKLVNSGKMGKSD	GAGAYCVASNEHGEVKSNEGSLKMLA	REDFRVRPTQVA	144	
Qy	144	LGGEAMVLECSPPGPEP	PSWRRDKELRIQDMPTITLHSDGNLI	IDPVDSDSGITQ	203	
Db	145	LGGEAMVLECSPPGPEP	PSWRRDKELRIQDMPTITLHSDGNLI	IDPVDSDSGITQ	204	
Qy	204	CVANNMGWGVSNPARLS	VFKPEQEPKMTVDVGAVALFCRVGDPQ	QITWKRKN	263	
Db	205	CVANNMGWGVSNPARLS	VFKPEQEPKMTVDVGAVALFCRVGDPQ	QITWKRKN	264	
Qy	264	EPMPVTRAYIAKMDNGL	RIERVQSDSGEYVCYARNPAGTLEASH	LRVQAPSPQTKPA	323	
Db	265	EPMPVTRAYIAKMDNGL	RIERVQSDSGEYVCYARNPAGTLEASH	LRVQAPSPQTKPA	324	
Qy	324	DQSPVAGGATFECTLVQ	QSPFAIYNSKGEQDQLLPSTVSADGR	ITKVSPTGLTILIEVR	383	
Db	325	DQSPVAGGATFECTLVQ	QSPFAIYNSKGEQDQLLPSTVSADGR	ITKVSPTGLTILIEVR	384	
Qy	384	QVDBGATVCAHNSAGSS	LKAAALATPEATKGRVQKSSKMGKQKQ	NSIKTLISAV	443	
Db	385	QVDBGATVCAHNSAGSS	LKAAALATPEATKGRVQKSSKMGKQKQ	NSIKTLISAV	415	

Qy 444 TGNTPAKPPPTIEHGHQNTLMVGSAILPQASGKPTGISWLRDGLPIDITDSRISQH 503
Db 416 TGNTPAKPPPTIEHGHQNTLMVGSAILPQASGKPTGISWLRDGLPIDITDSRISQH 475
Qy 504 STGSLHIALDKPDGTGYTCIAKNDGSGTWSASLVEDHTSNAQVFMPPDSNPFSSPT 563
Db 476 STGSLHIALDKPDGTGYTCIAKNDGSGTWSASLVEDHTSNAQVFMPPDSNPFSSPT 535
Qy 564 QPILVNVITDEVELHNAFSTSGAGPITGYIIQYSPDLQTFWNIPTDVASTETRIGL 623
Db 536 QPILVNVITDEVELHNAFSTSGAGPITGYIIQYSPDLQTFWNIPTDVASTETRIGL 595
Qy 624 KPSHSMFVIRAEKNGIGTFSVSSALVTTSPAAQVALSDKNKMDIAEKRLTSEQL 683
Db 596 KPSHSMFVIRAEKNGIGTFSVSSALVTTSPAAQVALSDKNKMDIAEKRLTSEQL 655
Qy 684 KLEEVKTINSTAVRLFKKKRLEELIDGYIIKWRGPPRTNDQTVVNTSPSTENYVSNL 743
Db 656 KLEEVKTINSTAVRLFKKKRLEELIDGYIIKWRGPPRTNDQTVVNTSPSTENYVSNL 715
Qy 744 MPPTNTEFFVPIHSGVSHHGAPSNMDVLAZAPPSLPEDVIRMLMLTLRLSKWA 803
Db 716 MPPTNTEFFVPIHSGVSHHGAPSNMDVLAZAPPSLPEDVIRMLMLTLRLSKWA 775
Qy 804 PKADGNGILGLGQIVIVGQAPNNRNITNERAASVTLFHTMTGKTIKVAASNGGV 863
Db 776 PKADGNGILGLGQIVIVGQAPNNRNITNERAASVTLFHTMTGKTIKVAASNGGV 835
Qy 864 GVSHTSEVIMNQDTLEKHLAQGNESFLGLNKSHPVIVVAILLIFFVILIANCY 923
Db 836 GVSHTSEVIMNQDTLEKHLAQGNESFLGLNKSHPVIVVAILLIFFVILIANCY 895
Qy 924 WNSNSKSDGDSFIKINDGVSMAINMDVQAQNPQPHNTAGRMTHNNKNGQALYS 983
Db 896 WNSNSKSDGDSFIKINDGVSMAINMDVQAQNPQPHNTAGRMTHNNKNGQALYS 955
Qy 984 LIPNAQGFNNCDYSGTMHRGSGSHHYHAQLGGPGNAMSTTGGVYHDDSPATT 1043
Db 956 LIPNAQGFNNCDYSGTMHRGSGSHHYHAQLGGPGNAMSTTGGVYHDDSPATT 1015
Qy 1044 LVLNQGPQAWLNDKMLRAPAMPTNPVPPPPARYADHTAGRSSRSASGRTGLNGLE 1103
Db 1016 LVLNQGPQAWLNDKMLRAPAMPTNPVPPPPARYADHTAGRSSRSASGRTGLNGLE 1075
Qy 1104 HRTSGSQRSQSPHTDYSVQLHSSDGTGSSKERTBERTPNTKLMDFIPPPSPPPP 1163
Db 1076 HRTSGSQRSQSPHTDYSVQLHSSDGTGSSKERTBERTPNTKLMDFIPPPSPPPP 1135
Qy 1164 GHRYD---TATRQLNRSGTREDTIDYSVQDGFARVDNARPTSRNKLGGRLPKG 1219
Db 1136 GHRYDIDTQATRQLNRSGTREDTIDYSVQDGFARVDNARPTSRNKLGGRLPKG 1195
Qy 1220 RDDSGQSSLMDDGGSSDADGSEGDVPRGQYKAVPRMIGSASTLHSCYCTNGTA 1279
Db 1196 RDDSGQSSLMDDGGSSDADGSEGDVPRGQYKAVPRMIGSASTLHSCYCTNGTA 1255
Qy 1280 QFRFSIPRNNIGTVTEQT 127
Db 1256 QFRFSIPRNNIGTVTEQT 1273

RESULT 2

T29548

hypothetical protein ZK377.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29548

R:Wan, M.; Hawkins, J.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid ZK377.

A:Reference number: Z20639

A:Accession: T29548

A:Status: preliminary; translated from G8/EMBL/DOB3

A:Molecule type: DNA

A:Residues: 1-874 (CDS)

A:Cross-references: EMBL:U88183; PIDN:AAB52657.1; GSPDB:GN00028; CESP:ZK377.2

A:Experimental source: strain Bristol N2; clone ZK377

C:Genetics:

A:Gene: CESP:ZK377.2

A:Map position: X

A:Introns: 91/2; 356/1; 452/1; 701/3; 746/3; 850/1

Query Match 67.5%; Score 4628; DB 2; Length 874;
Best Local Similarity 100.0%; Pred. No. 4.5e-245;
Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 MRSQKQNVQSIITLYLISAVTGNTPAKPPPTIEHGHQNTLMVGSAILPQASGKPTPG 483
Db 1 MRSQKQNVQSIITLYLISAVTGNTPAKPPPTIEHGHQNTLMVGSAILPQASGKPTPG 60
Qy 484 ISWLRDGLPIDITDSRISQHSQSLHIALDKPDGTGYTCIAKNDGSGTWSASLVEDH 543
Db 61 ISWLRDGLPIDITDSRISQHSQSLHIALDKPDGTGYTCIAKNDGSGTWSASLVEDH 120
Qy 544 TSNQVFMPPDSNPFSSPTQPILVNVITDEVELHNAFSTSGAGPITGYIIQYSPDLG 603
Db 121 TSNQVFMPPDSNPFSSPTQPILVNVITDEVELHNAFSTSGAGPITGYIIQYSPDLG 180
Qy 604 QTFWNIPTDVASTETRIGLKLPSHSMFVIRAEKNGIGTFSVSSALVTTSPAAQVALS 663
Db 181 QTFWNIPTDVASTETRIGLKLPSHSMFVIRAEKNGIGTFSVSSALVTTSPAAQVALS 240
Qy 664 DKMKNDMAIAEKRLTSEQLIKLEEVKTINSTAVRLFKKKRLEELIDGYIIKWRGPPRTN 723
Db 241 DKMKNDMAIAEKRLTSEQLIKLEEVKTINSTAVRLFKKKRLEELIDGYIIKWRGPPRTN 300
Qy 724 DNQTVVNTSPSTENYVSNLMFPPTTEFFVPIHSGVSHHGAPSNMDVLAZAPPSLP 783
Db 301 DNQTVVNTSPSTENYVSNLMFPPTTEFFVPIHSGVSHHGAPSNMDVLAZAPPSLP 360
Qy 784 PEDVIRMLMLTLRLSKWAPKADGNGILGLGQIVIVGQAPNNRNITNERAASVTLF 843
Db 361 PEDVIRMLMLTLRLSKWAPKADGNGILGLGQIVIVGQAPNNRNITNERAASVTLF 420
Qy 844 HLVTGMYTKIKVAASNGGVSGHSGTSEVIMNQDTLEKHLAQGNESFLGLNKSHPV 903
Db 421 HLVTGMYTKIKVAASNGGVSGHSGTSEVIMNQDTLEKHLAQGNESFLGLNKSHPV 480
Qy 904 VIVVAILLIFFVILIANCYWNSNSKSDGDSFIKINDGVSMAINMDVQAQNPQPHN 963
Db 481 VIVVAILLIFFVILIANCYWNSNSKSDGDSFIKINDGVSMAINMDVQAQNPQPHN 540
Qy 964 MYNTAGRMTHNNKNGQALYSLTPNAQGFNNCDYSGTMHRGSGSHHYHAQLGGPGNA 1023
Db 541 MYNTAGRMTHNNKNGQALYSLTPNAQGFNNCDYSGTMHRGSGSHHYHAQLGGPGNA 600
Qy 1024 MSTPYGQNHDDPSYATTTLVLSNQGPQAWLNDKMLRAPAMPTNPVPPPPARYADHTAG 1083
Db 601 MSTPYGQNHDDPSYATTTLVLSNQGPQAWLNDKMLRAPAMPTNPVPPPPARYADHTAG 660
Qy 1084 RRSRSASDGSRTGLNGLHHRSTSGQSDSPPHDYSVQLHSSDGTGSSKERTBERT 1143
Db 661 RRSRSASDGSRTGLNGLHHRSTSGQSDSPPHDYSVQLHSSDGTGSSKERTBERT 720
Qy 1144 PPNKTLMDIFIPPPSPPPGGHVDTATRQLNRSGTREDTIDYSVQDGFARVDNAR 1203
Db 721 PPNKTLMDIFIPPPSPPPGGHVDTATRQLNRSGTREDTIDYSVQDGFARVDNAR 780
Qy 1204 PITSRNKLGGRLPKGRDDSGQSSLMDDGGSSDADGSEGDVPRGQYKAVPRMIG 1263
Db 781 PITSRNKLGGRLPKGRDDSGQSSLMDDGGSSDADGSEGDVPRGQYKAVPRMIG 840
Qy 1264 SASTLHSCYCTNGTAQFRFSIPRNNIGTVTEQT 127
|||||

Db 841 SASTLAHSCGTGNGTAQRFSIPRNMGIIVTQBT 874

RESULT 3

T29549

hypothetical protein ZK377.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29549

R:Nhan, M.; Hawkins, J.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid ZK377.

A:Reference number: Z20639

A:Accession: T29549

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-423 <CHA>

C:Cross-references: EMBL:U88183; PIDN:AA852658.1; GSPDB:GN00028; CESP:ZK377.3

Experimental source: strain Bristol W2; clone ZK377

Genetics:

A:Gene: CESP:ZK377.3

A:Map position: X

A:Introns: 24/1; 142/3; 229/3; 284/2; 408/3

Query Match 32.5%; Score 2232; DB 2; Length 423;
 Best Local Similarity 100.0%; Pred. No. 1.1e-114;
 Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYLGFYHTHTHTHTYINFKIPNASHLAPIIEHPIDVWSRGSPATLNGAKPSTAKI 60

Db 1 MYLGFYHTHTHTHTYINFKIPNASHLAPIIEHPIDVWSRGSPATLNGAKPSTAKI 60

Qy 61 TWIDGQVITKMEQVNSHRIVLDTGSLFLKLVKNGKNGKSDAGATVCVSNHSESVKS 120

Db 61 TWIDGQVITKMEQVNSHRIVLDTGSLFLKLVKNGKNGKSDAGATVCVSNHSESVKS 120

Qy 121 NGSGLNMLREDRFRVPRVPTVQALGSEMAVLCSPPRGPEPVPVSWKRDDELRIQDMFR 180

Db 121 NGSGLNMLREDRFRVPRVPTVQALGSEMAVLCSPPRGPEPVPVSWKRDDELRIQDMFR 180

Qy 181 TYLSDGNLIIDPVRSDSGTQCVANNMGERVSNAPRLSVFEKPFQEPKMTVDVG 240

Db 181 TYLSDGNLIIDPVRSDSGTQCVANNMGERVSNAPRLSVFEKPFQEPKMTVDVG 240

Qy 241 AAVLFDCRVGTDPQPIITWRRKNEMPVTRAYIAKDNRLRIERVQPSDEGEVCTARNP 300

Db 241 AAVLFDCRVGTDPQPIITWRRKNEMPVTRAYIAKDNRLRIERVQPSDEGEVCTARNP 300

Qy 301 AGTLSEABLRVQAPPSFTQPKADQSVPGAGTATPCTLVGQSPATVSKKEQQQLLFP 360

Db 301 AGTLSEABLRVQAPPSFTQPKADQSVPGAGTATPCTLVGQSPATVSKKEQQQLLFP 360

Qy 361 SVSADGRKTVSPGTGLTIEEVQRVDEGAYCAGNAGSSLSKAAKATFETKGRVKK 420

Db 361 SVSADGRKTVSPGTGLTIEEVQRVDEGAYCAGNAGSSLSKAAKATFETKGRVKK 420

Qy 421 KSK 423

Db 421 KSK 423

RESULT 4

T30805

dutt1 protein - mouse

N:Alternate names: transmembrane receptor protein Robol homolog

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30805

R:Wu, M.C.; Lowe, N.; Pordhan, R.; Rabbitts, P.

submitted to the EMBL Data Library, July 1998

A:Description: The mouse homologue of human DUT1/H-robo gene: protein sequence and chr

A:Reference number: Z20879

A:Accession: T30805

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1612 <MUM>

A:Cross-references: EMBL:U17793; NID:el329712; PID:el329713; PIDN:CAA76850.1

A:Experimental source: brain

C:Genetics:

A:Gene: dutt1

A:Map position: 16

Query Match 21.9%; Score 1505.5; DB 2; Length 1612;
 Best Local Similarity 32.1%; Pred. No. 3.3e-74;
 Matches 402; Conservative 176; Mismatches 498; Indels 175; Gaps 2;

Qy 30 PVILIEHPIDVWSRGSPATLNGAK-PSTAKITWYKQPVITNKEQVNSHRIVLDTGSL 88

Db 29 PRIVEPVSOLIVSKGEPATLNGAKRGPTPTIEWYKGSERVETDKDPSHRMLPGSSL 88

Qy 89 FLILVKNKNGKSDAGATVCVSNHSESVKSNGSLKMLREDRFRVPRVPTVQALGGM 148

Db 89 FFLIRIVHKKSR-PDEGVITCVARNYLGAEVSNASLEVAIIRDFPQPSQVNVAVGEP 147

Qy 149 AVLECSPPRGPEPVPVSWKRDDELRIQDMFRPTVLTSDGNLIIDPVRSDSGTQCVANN 208

Db 148 ANWECQFPGPEPTITSKKQGLSDDDK-ERITIRG-KLMTITTKSDAGVTCVGTN 205

Qy 209 MWGERVSNAPRLSVFEKPFQEPKMTVDVGAAVDFCRVGTDPQPIITWRRKNEMPV 268

Db 206 MWGERSEVLTIVLERPSFKVPSGLNLTVDSEATYCAAGQVPPVPSWKRDDGELPA 265

Qy 269 TRAYIAKDNRLRIERVQPSDEGEVCTARNPAGTLEASHLRVQAPPSFTQPKADQSV 328

Db 266 SR-TEKRDHILAKRYTAGDSGTVCARMVMGKASASLTIVQEPFVPPVPAQVPA 324

Qy 329 AGGTATFCTLVGQSPATVSKKEQQQLLFPST-VSADGRKTVSPGTGLTIEEVQRV 386

Db 325 LGRVITQVCEATGMPQPAIFWRRKSGQLFP-STQPPSSSSFSVSGTGLTITWQRSD 383

Qy 387 EGATVCAGNAGSSLSKAAKATFETKGRVKKKSKMGQKQKRVSIITKLISAVTGN 446

Db 384 VGYITQTLNAGSIIITAYLE-----VTDV 409

Qy 447 TPAPPPPTIEHGHQNTLMVSGSALLPQASGAPRTGISLRDLGDLIDITDSRIQSHTG 506

Db 410 IADRPPVPRVQGVNQTVAVDGLTLLSCVATGSPATILNRKQGLVLTSDSRIRQLSEG 469

Qy 507 SLHIAADLKPQGVTTCTIAKNNGESTWSASLTVEDETSNAQFVRMPDNPFPSSPTQPI 566

Db 470 VLQIRAKLQGTGRITCTASTPSGATVSAIEVQEGVQVPPFRPTDPLILPSAPKPE 529

Qy 567 IWNVTITVELHWNAPSTSGAGPITGYLIQYSPDLQGTWFINPDTVASTETIRKLGAPS 626

Db 530 VTDVSKNTVLISQNTLNLSGATP-TSYLBAFSAHSGSSQTAENVTETPAKGLAPN 588

Qy 627 HSMFVIRAEKNGIGTSP-VSSALVTTSKPAQVALSKNKKMDMAIEKRLTSQQLIK 685

Db 589 AITLFLVRAANAGISDPSQISDPVKTQVPPTSCQVDEKQ-----VQREGLWVLEH 641

Qy 686 EEVKTINSTAVRLFWKKKLELIDGTIYKRGPPRTN-DNQIV-NVTSPTSTMTVYSN 742

Db 642 NHTPLSSSSVEVNTVDQSQYIQGYKILYRPGASHGESEMLFVEVRPTTKNSVIVPI 701

Qy 743 LMFPTNYEYFVIPHSGVHSIHGAPSNMVLTAEPAPSLPEDVIRML-NLTLTIS 800

Db 702 LRKGVNYELKARP-----PNEQGDSEIKFPAKLEAPASPPSVTVSKNDGNTALITV 758

Qy 801 WKAPFADGICLQAGQIVVQAPNNRNITNEAASVTLPELVKTYKIRVAASN 860

Db 759 WQPPEDTQNMVQVETKWLQNETYHINKTVDSGTSVVIVLGVIRYSEVAASGT 818

Qy 861 GGVGV-----SGTSEVMNQDLEHAAQJNESFLYGLNKSHSPVIVIAI 910

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	5
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

Qy 912 IIPVILIIATCTWNRNSDKGRSFIKINDGSVHMSN-----NLN-----DV 955
 Db 1101 TIIVIVIVAVPCTRTTSHQKRAACKSVMSHKTGSKDKVDPDLNIIHRELKPI 1160
 Qy 956 AQNPWNPNMTAGRTMNNRNGALYSLTPT-NAQFPNCCDSTGMRPOSEHHY 1013
 Db 1161 DKSPDPNMTD-----TPIPNISQ-----DITPDNSNG-----SNIGQRNRSRGHE 1204
 Qy 1014 AQLTGQPGMNAITFYNGYHEDDPSYATITLFLSNQCPAWLMDKMLRAPAMPTNVPVPE 1073
 Db 1205 SE-----DSMTLAGRR-----GMRPKM-----MPDQPPQ 1233
 Qy 1074 -----PARTADTAGRRSSRSRSDGRTLMLGLHRTSGSQRSSPDEHT----- 1119
 Db 1234 VISAPHISLDNPHHPSGSLASPTSR-----LHQVSPWPVGTSMHSRANSTSVR 1289
 1120 -----YSTQLHSDGT-----GSKERTGERRPT-----PKTLMDFIPP 1156
 1290 NTPSSDTPMSSGQPCADHQDQDSSSGAYLGSAGEEDAAQSLPTAVRPSHPLKSAFV- 1348
 Qy 1157 PSNPPTGGPHGV-DATRQLNRGSTPRDRTD-----SVSDGARFVWVARPTNRNGLG 1213
 Db 1349 -----AVPAASADDTPLTPTLLTQAPSHVPSVKATSTGLGR-----TRPP----- 1393
 Qy 1214 RPLMGKRGDSQRSLLMDGSSSEAD 1241
 Db 1394 MPVVPSPAPVQETITMLEDSSESTPD 1461

RESULT 9

A53449

plasmacytoma-associated neuronal glycoprotein PANG - mouse

C:Species: Mus musculus (house mouse)

C:Date: 25-Aug-1995 #sequence_revision 25; Aug-1995 #text_change 24-Sep-1999

C:Accession: A53449

R:Connelly, M.A.; Grady, R.C.; Mushinski, J.F.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 91, 1337-1341, 1994

A:Title: PANG, a gene encoding a neuronal glycoprotein, is ectopically activated by inta

A:Reference number: A53449; MUID:94151325

A:Accession: A53449

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1028 <CON>

A:Cross-references: GB:101991; NID:g200059; PIDN:AA17403.1; PID:g200057

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

Keywords: glycoprotein

Query Match 9.04; Score 6.6; DB 2; Length 1028;
 Best Local Similarity 24.94; Pred. No. 6.3e26;
 Matches 244; Conservative 135; Mismatches 421; Indels 178; Gaps 35;

Qy 30 PVIIIEHP-----IDVVSRSGPSATLNCAGKPS-TAKITWYKDGQVITNKEQVNSHRLVDT 85
 Db 26 PVIEKPSNISIIVDSQDKITLNCAGRNPSYRQWLNGSDIDTSLD-----HRTKLG 81
 Qy 86 GSLFLAYVNSGKNGSDAGAYCVASNEHGVNSGSLKMLNREDFVRPR-TQVAL 144
 Db 82 GNLIVNPS-----RNDTOSYQCFATNSLGTIVSRKALQPAYI-ENFATRGSRSTSVR 135
 Qy 145 GGEAVLSCSPGPFPEPVVWSRDKDELRI-QMPTYLHSDGLIDPVDSDSGTQ 203
 Db 136 GGGVLLCPGPPHSGELSTAWFYNEYSFVDESKRLVSQETGLHIAKVEPSDNGNT 195
 Qy 204 CVANNNGER-VSNPAKLSVE-----RKPE-QEPKDMVWGVANGLFDCRWGDP 253
 Db 196 CVVTSIVNTRVLSGPTPLVLRSDGVMSGETPKLVQVFPETLPAANGSVLSCPALGN 255
 Qy 254 QPQITWKRKN-EPMPVTRAIKDNRLRIERQVSDGSEYCYANRPAQTLBSAHLRV 312
 Db 256 VPQINRRSDGMPF-NKIKLRKNGMLIQNQBDRTSGEIGLNSRGNVARGTLY 314

Qy 313 QAPSPQTQKADQSPVAGGATTAPECTLVQGPSPAYFMSKEQGLDLPFFTSVADGRTKVS 372
 Db 315 YAKPTIQLLDVETIAVEDSLTWEBCASRKKPSYRWLNGDALVL-----EERIQIE 367
 Qy 373 PTGTLTIEEVQVDEGATVCAGMNSAGSSLKAAKAT----- 410
 Db 368 -NAGLITNMLVTDVSGQPCIAENRGLIYSSLAELVAVASAPDRSPNMKKVQVQVGLS 426
 Qy 411 -----PETGRVQ-----KKSKMGKQKQK-----NVQSIK-----YLISAV----- 443
 Db 427 VILDCPRASPRALSPFKWKGDMVREQARVFLNDGGLIMNVITADAGTITCTAENQFG 486
 Qy 444 -----TGNTPAKPPPTIEHGQNQTLVGSALLPCQASGKPTPTGI-SMLRDLGLDITD 497
 Db 487 KANGTLLVTEPTRLILAPSMDVAVGESVILPCQVQDPLLDIMFANTFNGALDFKK 546
 Qy 498 -----SRISOHSTGSLRIADLKKPTDGYTTCIAKNDEGSTVASLTVEDHTNSQFVYM 552
 Db 547 DGSHEKVGSGSSGDLIRNIQLKHSQYICMVTQGVDSVSAELIV----- 595
 Qy 553 PDPSNFPSPQIPIVNVITDEVELHNPASTSGAGPTIIGYTSP-DLG-QTWNPFI 610
 Db 596 -----GSPGPPENKVDIEITDTIAQLSW-TESTDSHPVISYAVQARTPSVGQSVKTPV 550
 Qy 611 DVFAYTEY-RIGLQKPSHSYMFVIRAEENKIGCTPSVSSALVTTSKPAQVSLDKMK 668
 Db 651 EVIDGKTHTATVVELNPMVVEYEFRIAVANSKIGGEPSELSEKRVTEAAPETAPSEVS- 708
 Qy 669 DMAIAEKRLISBQLIKLEEVKTINSTAVRLPKKRLLELD---GYIKWGGPFRPD 724
 Db 709 -----GGGSGSLVITWDVPP-----EELQGGGFTVVFAPRLGVITV 748
 Qy 725 NQVNVNTPSPSTENYVSN-LMPFTNTEPPIVTHSGVSHGAPSNMOLVTAAPSL 782
 Db 749 IQTV-VTSPDPRVYFRNESIVPSPFTEGVGVYN-----NKGCGPSPVTTVSAAEPTV 804
 Qy 783 PPEDVRIMLMLTLIRSNKAPKADGINGILKQPIVI-VQAPNNRNNITNTRAAVS 840
 Db 805 APHSISABLSLSSEIIVSWMTIPWLSNGLHLYEYRVNNGNGEESRKKVAGNQTS 864
 Qy 841 TLFLHVTGMTIKIRVARSNGGVSHGT-----SEVIMQD 877
 Db 865 VLRLGSLNLAITVAVRAYNSAGAGPFSATVNTTKTPSPGPNVNNATDTVLNME 924
 Qy 878 TLEKHLAQGEENSLTG 955
 Db 925 QVK-----AMENESEVTG 937

RESULT 10

T13822

frazzled gene protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T13822

R:Kolodziej, P.A.; Timpe, L.; Mitchell, K.J.; Goodman, C.S.; Fried, S.; Jan, L.Y.; Ja Cell 87, 197-204, 1995

A:Title: Frazzled encodes a Drosophila member of the DCC immunoglobulin subfamily and

A:Reference number: T13780; MUID:97015076

A:Accession: T13822

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1375 <COL>

A:Cross-references: EMBL:U71001; NID:g1621114; PID:g1621115; PIDN:AA047314.1

C:Genetics:

A:Gene: frazzled

A:Map position: 2

C:Function:

A:Description: may function in vivo as a receptor or component of a receptor mediating

Query Match 8.84; Score 507; DB 2; Length 1375;

Best Local Similarity 21.64; Pred. No. 3e-25;

Matches 304; Conservative 182; Mismatches 488; Indels 434; Gaps 60;

Qy 36 PIDWVSRGSPATLNC-----GAK-----PSTAKITW-KDGPVITNKQVNSHRI 81
 Db 41 PQDAVPGDSVLLQAGTASTGRGGSKMLPSSVSIWRPQDQDLVIGD-----TFRT 97
 Qy 82 VLDTGSLFLKLVGKMGKSDAGATYCVASNEH-GEVSNDEGSKLAM-----LREDPRVR 137
 Db 98 QLKNGSLYSSYEENR-----GTGATYQLLABRGVGLRPAVAVIWPOLQMOF----- 150
 Qy 138 PRIVQALGEMAVELECS-PPRFPPFPV-----VSWRKDKELRQDMRITLHSDGNLIID 193
 Db 151 -LETTLLPQQTAYFRCLMGDAWQEGVHSVQWKDLDLPL-DLKLAWVLPGLALEIDE 208
 Qy 194 VDRSDGTTQCVANMNGERVSNPARLSVE-----KPKFEPKMDTVGVGA 243
 Db 209 VGPSDRGSGTQCVNTGSSSLRSKTNLKKPSQGVKSVAPSPLVGPSKTVREGDITV 268
 Qy 244 LDFCRVQDPQOQITWKKNEPMPVTRAIKDNK-----GLRIEVRQPSDEGTVCY 296
 Db 269 PLDCVANGVPKQIKWLRNGMDLD-----FNDLSRPSIVGTGSLISSAEDISGNTQCR 324
 Qy 297 ARNPAGTLESAHLRVAQPPSPQTKPADQVPAQTATPECTLVGQSPATVSKSGQD 356
 Db 325 ASNTVDSLAQAVVQVQPPKFIKPKDTTAEHDEPELCKDIWGPVKWIKMG--D 382
 Qy 357 LLFSPS-TYS-ADGRITVSPGTLLTEEVQVDGAVYCAHNSGSLSKAL----- 407
 Db 383 LITPDYQVLDGK-----NLKTLGLLNSDAGMPCVGTWAGSVHAARLRVVPQD 435
 Qy 408 -----KATFET----- 413
 Db 436 SPDQPSVPHVGGKPLDSGLQARLPSQPRDLVAQIKSVFVLSWEPFLQAGDVYITV 495
 Qy 414 -----KGRVKKKSKMGKQKQVWSIIKYLNAVGTMPAKPPTIEHGWQVLMWG 467
 Db 495 YTKMNSERQDQWVASHDDQVNIQSL-----PQRTQFVREANTFSG 541
 Qy 468 SSALPLQCAQSGPTGIS-----WLRDLPI-----DITDSRI-----SQHST 505
 Db 542 GASAPLEVTQSPVNIAGPPRFNFGTASKEVITVKEEPTVINGELIATVYISGDS 601
 Qy 506 GSHLHDLKPKDGT-----VTTCAKNEDEGSTVASLIVDEHNSNQVFR 551
 Db 602 G-----ADLTDESTALEANVLELRPBTIVISVVPFRNKGMDSGSAEIRKVFST 652
 Qy 552 MPDPSVPSPPSQPIQVWVIT-DTEVELWNPASTGA-GPITGIIQY-----SPOLQGT 605
 Db 653 -----PSEFPNNVLTGVSSTIVBEWPAEEDRNGQITGKIRTKAFKDAQVST 705
 Qy 606 -----WFNPIDVASTYERI----- 620
 Db 706 PANIRTFELSNLDKNAETQVKIAMTVNGSGPPTENNRANTLENLDGTQVPGKPIWISI 765
 Qy 621 -----KGLPSHSTMF 631
 Db 766 HPGANNIALEWGPQHPPEIKIRVYLVGWRGCPDENTIELKERTYHILNLSNDITV 825
 Qy 632 VIRAEKNEGTGSEVSSALVTISKPAQVSLSDKNDKMAIAEKLSEQLIKLEBKVTI 691
 Db 826 SLARNVKGDGPPIIDNKTREDEP-----VDAPITLEVVPGLAI-----TM 868
 Qy 692 NSTAVLEW-----KARKLELIDG-YIKRWGPPRTNDQVYVNTSPSTENTVSNLMPT 747
 Db 869 SSSSIVYTDMLNKKQVYDNRHTYVSYGITGSKRYTRHTTD-----LMCMNDLRPNT 925
 Qy 748 NYEZFVPTHSVGHSH--APNSMDVL-----TAAPSLPPEVDRIAM--LNLTLRLSN 801
 Db 926 QYEF-----AKVTVGREGSSMSVLSNTYQVVPVPPRETVRLDDNDPQIVQV 978
 Qy 802 KAPKADGINGILKQFVIVQAPNNRNITINERASVTLF-----HLVTGMYTKIRVA 858
 Db 979 IPPK--HTLQGITGNIYITTDITKRRDWSVEAFGEETMLNPLMKPTTYTTFVQKAR 1036

Qy 859 SNGGVG-----VSHGTSEVLMNQ--DTLEKHLAAQGENESFLYGLINKSHVPIVIVA 909
 Db 1037 TTGANNAPPSALVSTTSAATVQBPETIAGI-----DNEKLLY-----IIIAA 1082
 Qy 910 ILIIFVVI-----IIATCYWRNSNSGDKRSFIKINDGSVHMSNMLWDAQNPNMNYN 956
 Db 1083 TAVVLVLLVGLLILCRRKPSQSPETHKSTQKNVGV-----PKPDLMI 1128
 Qy 967 TAGRMTNNNNQALYSLTPAQDFNNCCDYSTNHRPQSEHYHTAQLTGCGCANMST 1026
 Db 1129 HDHQMLKLNID-KGLHTVTPVCSGASS-----SGALTLPVSHSEYETVETVPVGHVNS 1183
 Qy 1027 FYGNTDHPDPSPIATTLVLSNQPAWLDKMLRAPAMNTNPVPEPPARYADTAGRS 1086
 Db 1184 LDKRST-----VGTMTTS-----MNGTMR-----PQTP-----RT 1210
 Qy 1087 BSSRASDGRGLNGLHRTSGSQSDS-----PPHT-----DVSTQL-HSSDGTGSSKER 1137
 Db 1211 QVSHQWRSMTMEAGLSQSLTQPSNSMAQTPEHPGYGDYDNCNAGNAAAGNCVSTI 1270
 Qy 1138 TGERRTPNKTLMDFPPIPPSPNPPGG 1165
 Db 1271 BSKRGHP-----LKSFSVP-----GPPPTGG 1292

RESULT 11

T08851

Down syndrome cell adhesion protein 1 - human (fragment)

N:Alternate names: Down syndrome cell adhesion molecule

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C:Accession: T08851

R:Yamakawa, K.; Buo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.: Submitted to the EMBL Data Library, September 1997

A:Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down

A:Reference number: Z16495

A:Accession: T08851

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1896 <IAB>

A:Cross-references: EMBL:AF023449; NID:g169765; PID:g169766

A:Experimental source: brain; developmental stage: 14 weeks; fetal

C:Genetics:

A:Gene: DSCAM

A:Map position: 21q22

A>Note: derived from alternately-spliced mRNA

C:Function:

A:Description: involved in nervous system development

C:Keywords: alternative splicing

Query Match 8.7%; Score 598; DB 2; Length 1896;
 Best Local Similarity 25.24; Pred. No. 1.4e-24;
 Matches 230; Conservative 131; Mismatches 365; Indels 188; Gaps 41;

Qy 16 YINDKIPNASHLAPVIEHP-----DVVSRGSPATLNCGRPS-TAKITWYKDG 66
 Db 372 FYRKDL-SAQDTQVQVLEDTPKIISAFSEKVSYPAPVSLMNCVNGTPLPTITLDD 430
 Qy 67 QPVIITNKQVNSHRI-----VLDTGSLF-LKLVYKMGKMGKSDAGATYCVASNEHGEVSKNE 122
 Db 431 DPLKGL-----GSHRISQMITSEGVNSVYLNIS-----SQVRGGVYRCTANNAGV---- 479
 Qy 123 GSKLALMREDPR-----VRP-RTVQALGEMAVELECSPPRGPFVPSVNRKDKDELRIQD 177
 Db 480 -----LQARINVRGPASIRPMKNITAIAGRTYIHC-RIGTPTYSIKWNKSNLPLFNH 534
 Qy 178 MPRTLSDGMLIIPVDKS-DSTGTQCVANMNGERVSN-----PARLSVFEPK 226
 Db 535 R-QAFENGTNKLSDVQVKEDEGTC--NLVQPOLSTQSQVHTVYVPPPIQPEFP 591
 Qy 228 KFEDEPKMDTVGVGAFLDC-RVTGDQPIQITWKKNEPMPVTRAIKDNK-----RGLR 281

Db 592 RF-----SIQGRVIFPCVWSGDLPIITWQKGRGPIPSGLG-VTIINDITSSLR 641

Qy 262 IERVQPSDEEYTCYARNPAGTLEASRLVQAPSPQTKPAQDSVAGGTATFCTLVG 341

Db 642 ISNLSLHMGNTTCIARNEAAVEHQSLVVRPKPVVQPPQDGI7KRAVLNCSAG 701

Qy 342 QPSAFPM--SKGQDQLFPSTYSADGRTKVSPTGLTIEVRQVDDEGATYCAAGNSG 399

Db 702 YPVPTLVNKSAGVAGVQFQF--IALNGRIQLVSGSLKLVVDESGVLYCKVNSVDG 759

Qy 400 SLSLKAALKATFETKGVQKKKSMKGQKQKNVSIKILISAVTGMTPAKPPITIEHG 459

Db 760 ADVSKS-----MYLTWKI-----PAMITST 779

Qy 460 QNQTLMV-GSSAILPQASGKPTPGISWLRDCLPID-----ITDSRISQHSSTGLA 511

Db 780 PNTIIATQQCKKMSCTAHGKPIIVRWKEDRIINPMARYLSTKVEGEVISTLIQL 839

Qy 512 DLAKPDTGYTICIANKEGDESTWASLVEDHTSNAQVMPMDSPNFPSTPTQIIVNVT 571

Db 840 PTWVEDSGFPTSCAINSYSGEDGIIQLVQV-----PPDPEIEIKDKWK 883

Qy 572 DTEVELHNPASTSGAGPTIGYIIQYSPDLQGTWNI-----PDYASTETRIKGL 624

Db 884 ARTITLWMT-GFDGNSPTIGYDIE--CKNKSQSWDSQAORTKVDSPQLNSAT-----TIDH 937

Qy 625 PSHSTMYFIRAENEGIGTSPVSALVTYSKPAQVALDSKMKMMAIAEKRLTSEQLIK 684

Db 938 PSSTYSIRWAKNN--IGKSEPSNELTITADEAPDG-----PPQEVH 978

Qy 685 LBEVKINSTALFWK--KRLKE-LIDGYIKWNGPPTINDQIVNV-----TSPSTEN 737

Db 979 LE--PISSQSIRVWAKPKHLQNGIIRCTQIYR-EYSTGNGQPNVISVDTSDGSEV 1034

Qy 738 YVNSLMPPTNYEFVPIYHSGVSHGAPNSMDVLA--EAPSPDPEDVIRKMLMT 795

Db 1035 TYLDNANKPTQIGVLVQACNRA-----GTCPSSIEITTLLEDPVSTPPENQVAITSPE 1089

Qy 796 PLTISWAPKADGINSIKGQIVIVQAPNN-----RMITPNERAASVTLFEVGTMY 851

Db 1090 SISISNYSKALNGLQGFVRIWMLMDGELGEIKMTTQ--PSLEDGLKRYTNY 1147

Qy 852 KIRVAASNGGVG 865

Db 1148 SIQLAFTRAGDV 1161

RESULT 12
JU0094
F11 protein precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 31-Jan-2000
C:Accession: JU0094
R:Brummendorf, T.; Wolff, J.M.; Frank, R.; Rathjen, F.G.
Neuron 2, 1351-1361, 1989
A:Title: Neural cell recognition molecule F11; homology with fibronectin type III and in
A:Reference number: JU0094; MUID:90180453
A:Accession: JU0094
A:Molecule type: mRNA
A:Residues: 1-1010 <BRD>
A:Cross-references: GB:K14877; NID:g107884; PIDN:CA33018.1; PID:g63385
A:Note: the carboxy- and hydrophobic stretch is compatible with the consensus motif for C
A:Note: F11 comprises six domains related to the immunoglobulin domain type C and four
CAM
C:Comment: F11 is a chick neural cell surface-associated glycoprotein implicated in neur
C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology
C:Keywords: blocked carboxyl end; glycoprotein; immunoglobulin; phosphatidylinositol linked
F1-20/Domain: signal sequence #status predicted <CAT>
F:21-984/Product: protein F11 #status predicted <CAT>
F:247-303/Domain: immunoglobulin homology <MD>
F:985-1010/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F:200,249,329,448,464,485,512,582,621,924/Binding site: carbohydrate (Asn) (covalent) #

F:984/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature

Query Match 8.7%; Score 597; DB 2; Length 1010;
Best Local Similarity 23.6%; Pred. No. 6.7e-25;
Matches 268; Conservative 140; Mismatches 364; Indels 364; Gaps 47;

Qy 6 FTYHTHTHTHTY-----NFKIPNASNAPVIEHPIDVVV-----SRGSPATLMCGAKP 55

Db 3 FFIHSHTLCTPCVADSTHFSRQNG-KYGPVFEQGTIDITYPESSDSQVSMQCAR- 60

Qy 56 STAKITWYKDGQVITNKQVNSHRVL--DTGSLFLKYNVSGKNGKSDAGATYCVASN 113

Db 61 -----AVFPPTTKWKLNMWIDLTIKDRYSNMGVGLVSNPEKSDAGKTYCVSN 110

Qy 114 ERGEVSNNGSLKMLM-----REDFRVRPTVQALGEMAVLPCSPRPGFPFV-VSNR 167

Db 111 IFGTVSSBATLSFGYLDPPFPEEHVEYKVR-----GVGAVLCEPHYTPDOLSTWML 165

Qy 168 KDKLELIQ-DMPRTITLHSDGLIIPVDSRSGTQCVANNWGVSRVSA-RLSVFEK 225

Db 166 LNEFPFTIALARRFVSQTNMGVLYANVASDKNGSCF-----VSSPITSVFSK 217

Qy 226 ----PKFQK-----PKMTVD-----VGAVALFDCRVTDGPOLITWKRKNEMPVT- 269

Db 218 FIPLIQLADRAKVPADIKVKPKDYALGLQNVLTCLFALGNVPKRWKSLTEPMPATA 277

Qy 270 -----RAYI-----AKD----- 276

Db 278 EISMGGAVLKIFNIQYEDBGLYECBAENYKGDKHQARYVYQASPEWHIMTEKDG 337

Qy 277 -----NRG-LRIERVQPSDEEYTCYARNPAGTLEASRL 310

Db 338 DLYWPCVATGAPIPTIRKLVNWSVFKRELRIQSLTFEDAGWYQCIENNAHITANREL 397

Qy 311 RVQV-PPSPQIKPADQSVPA--GGATETCTVIGQPSQVATYSGQDQLFPSTVSADG 367

Db 398 KIVASPTPELWPKMKKILAAKGRVIECKPAKPKKSWK-GTELVWGS----- 450

Qy 368 RTKVSPGTLTIEVRQVDDEGATYCAAGNSGSLSKAALKATFETKGVQKKKSMKGK 427

Db 451 RIHWDSGLSIIIVNTKLDGRATFCAENNGKRGANSTGLVMTETR----- 497

Qy 428 KQKNVYSIKTILISAVTGMTPAKPPITIEHGHNQMLWSSAILPQASGKPTPGIS- 485

Db 498 -----TLAPLVNDVTYGENATQCIASHDPTLDLPFI 530

Qy 486 WLRQGLIDIT-----DSRISQHSSTGLAIDPKLTGTYTICAKNDEGSTWASRL 539

Db 531 WSLNGVLPIDKEKEHEHYRWKNSKNGELLKVLQKHAGRTCTAGTIVNDSASALV 590

Qy 540 VEDHTSNAQVMPMDSPNFPSTPTQIIVNVTDEVELHNPASTSGAGPTIGYIIQYTS 599

Db 591 -----VWGP-----PGPGGIRIEEDRIWALVWSR-GTDNHSPIKTYT-SK 633

Qy 600 PDLQGTWNIPTYVASTE-----YRIKGLPSHSTMYFIRAENEGIGTSPVSSA 649

Db 634 TFLSEWMD-----AKTSPDIEGNNMESARVIDLIPWMEYEFPIATNLTGTGSPMSQ 686

Qy 650 LVVTSKPAQVALSD-----KKNMDMAIAEK-----RL 677

Db 689 RIRTBGAPVWAPSDVGGGSGNRELITWMLPSREHYGNMGFTIVAPKPKGKERNRV 748

Qy 678 T-----SQLIKLE-----EV 688

Db 749 TVTNPEIGRTYKDESMPTSTQYKVVAFNSKGDGPFSLTAVTISAQDAPIETVDVSV 808

Qy 689 KTNSTAVLWPKMKRLLELDIYIK-WRGPPTNDNQVNYVSTSPSTNY--VYNSLMP 745

Db 809 KVLSSSEISLVN-HVETKSVSGYQIRVWAAHDKAAARQVQ-----SNQYETKLENLK 864

Qy 746 PTNYEFVPIYHSGVSHGAPNSMDVLAAPPSPDPEDVIRKMLNLTTL-----I 799

QY	4	LGETHYTHHTYINFDKIPNAPSLAPVIEHPIDVGVSRGS-----PATLMGAKPSTAK	59
Db	19	LGETHYTHHTGHWGSEEDK-----GGPIPEZGPNTPIYEESECKVSLACARASFP	73
QY	60	ITWKY-----DQGVITNKEQVNSHRIYLDGSGFLKLVNSGKMGDSAGAYCVASNEH	113
Db	74	V-YKRWNNQSDVLIN-----DRYSWGVGMIV-----NPDFKDKDGITVCLASNN	120
QY	116	GEVYSNGLSKLALM-----REDFVRVPRTYQALGGMVNLCESPPGPEPPV-YKSRKD	170

contactin precursor - chicken
N:Alternate names: 130K glycoprotein
C:Species: Gallus gallus (chicken)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 21-Jan-2000
C:Accession: S01998
R:Ranscht, B.; Dours, M.T.
J. Cell Biol. 107, 1561-1573, 1988
A:Title: Sequence of contactin, a 130-kD glycoprotein concentrated in areas of interneu-

Qy	6	PTATHTHTIYI-----NFQIPNASHAPVIEHDIVVY-----SRGSATLNCAGP	55
Db	3	FFSHILVLCIFPCVADSTHFSEZGN-KYGVFPVEQEPIDITYPESSDQVSMNCAR-	60
Qy	56	STAKITWIKDQGPVITKNQWHSRVLV--DGSLSFLFKVSGKNDSDAGATYCVAN	113
Db	61	-----AVFPPTTKMLNWDILTKDRTSMVGGRLVISNPKSDKSGATYCVSN	116
Qy	114	EHGEVSGNKGSLKLALE-----REDFRVRPTVQALGEMAVLECSPPGPPV-VSMR	167
Db	111	IFQVSSSEATLSFGTLDPPFPEEYEVKRV-----GGAVALCEPPTHYPODLSTWKL	165
Qy	168	KDKKELRIQ-OMPPTITLSDGNLIIDVDSRSDGTQCVANNMVGERSVNP-RLSVFK	225
Db	166	LNBPFFVITALLRRKVFQSGNGLYIANEASDGMKSYCF-----VSSPSITKSVFK	217
Qy	226	-----PKFEQE-----PKDMTVD-----VGAVALFCRVTDGPQOPITTKRKNEMPMTV	269
Db	218	FIPLIQADRAKVPADIKVVKFDYIALLGQNTYLCEPALGNVPELRLWSKYLEMPATA	277
Qy	270	-----RAYT-----AKQ-----	276
Db	278	EISMAGVALKFNFIQYEDGELQCEAEYNKGGKQKQRYVQASPEWHIHDITKEIGS	337
Qy	277	-----NRG-LRIERVQPSDEBQYVCTARNAGTLEASAL	310
Db	338	DLTYWCVARGAPITPIRLKNGVSSFRKGEIRIGLTIEDAGMTQCTABAGHIIYANAL	397
Qy	311	RVQA-PPSPQTKPADQSVPA--GGTATFECTITGVGSPATYFWSGQDQLPFSYVSADG	367
Db	398	KIVASPPPTFELNPKMKKILAAAGGRVILIECKRAAPKPFKSWSK-GTELLWGS-----	450
Qy	368	RTKVSPGTLTIEEVQVDEGATVCAGNKSAGSSLSKALKATPETKGVQKKSKKMGKQ	427
Db	451	RIEHWDSGLIEIIVTLKDBGRTYFCABNNGRKNASTGVLEIETRR-----	497
Qy	428	QRHNVQSLIKTLISAVGTMPARPPTITIEHQHNTQMLVGSSALILPQASCKPTPGIS-	485
Db	498	-----TLAPLNVDTVGENATWQCIASHPDLDLTFI	530
Qy	486	WLRLDGLPIDIT-----DSKISQHSSTLHIAHLKPKDTPYVTCIAKNEGDESTWSASLT	539
Db	531	WSLNGVFLDFCKEHEHYERNWIKNSGELLIKNVQLRHAGRTCTAQTVDNSSASADLV	590
Qy	540	VEDHTSNAGFVPMFPMSPNPSPTQPIIVNVVTDIEVELHWNAPSTSGAPITGIYIQQYS	599
Db	591	-----YRGCP-----PGPPGGIRIEIRDTAVALTNSR-GTWNESPISKYTIQ-SK	633
Qy	600	PDLGQTFWNIPIDYASTE-----YRIKLKPSHSYMFVIRAEKNGIGTSPVSSA	649
Db	634	FTLSEEDWD-----ARTKPSDIEGNMESARVILDPNWEYEFRIIANTVLTGCTGSPMSPQ	688
Qy	650	LWTSKPAQVALSD-----KKNMDMAIEAK-----RL	677

Search completed: January 22, 2001, 12:26:10
Job time: 2047 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

ON protein - protein search, using sw model

Run on: January 22, 2001, 12:27:59 ; Search time 162.41 Seconds
(without alignments)
257.899 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYTLGFTHTHTHTYINFD.....TAQRFRSIPRNGNIVTQBT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
	1	674	9.8	1493	1	NEOL_MOUSE
	2	642	9.4	1377	1	NEOL_RAT
	3	639.5	9.3	1461	1	NEOL_HUMAN
	4	632.5	9.2	1443	1	NEOL_CHICK
	5	598	8.7	2012	1	DSCA_HUMAN
	6	597	8.7	1010	1	CONT_CHICK
	7	595	8.7	1020	1	CONT_MOUSE
	8	593	8.6	1018	1	CONT_HUMAN
	9	585	8.5	1040	1	AXOL_RAT
	10	583.5	8.5	1257	1	CAML_HUMAN
	11	582.5	8.5	1040	1	AXOL_HUMAN
	12	567.5	8.3	1260	1	CAML_MOUSE
	13	562.5	8.2	1036	1	AXOL_CHICK
	14	561	8.2	1284	1	NRCA_CHICK
	15	557	8.1	1259	1	CAML_RAT
	16	554	8.1	1239	1	NRG_DROME
	17	554	8.1	1447	1	DCO_HUMAN
	18	552	8.0	1447	1	DCO_MOUSE
	19	526	7.7	1897	1	PTPF_HUMAN
	20	509.5	7.4	1912	1	PTPD_HUMAN
	21	467	6.8	2029	1	LAR_DROME
	22	462.5	6.7	1092	1	NCA2_XENLA
	23	459.5	6.7	1088	1	NCA1_XENLA
	24	455.5	6.6	1091	1	NCA1_CHICK
	25	455	6.6	1266	1	NGCA_CHICK
	26	441.5	6.4	1070	1	PTK7_HUMAN
	27	435.5	6.3	1115	1	NCA1_MOUSE
	28	433.5	6.3	853	1	NCA1_BOVIN
	29	433	6.3	858	1	NCA1_RAT
	30	432	6.3	848	1	NCA1_HUMAN
	31	432	6.3	1051	1	PTK7_CHICK
	32	429	6.3	761	1	NCA2_HUMAN
	33	427.5	6.2	725	1	NCA2_MOUSE

34	425.5	6.2	3707	1	PGBM_MOUSE	Q05793	mus musculus
35	423.5	6.2	837	1	NCM2_HUMAN	O15394	homo sapien
36	422	6.2	811	1	FS22_DROME	P34083	drosophila
37	422	6.2	4393	1	PGBM_HUMAN	P98160	homo sapien
38	409.5	6.0	837	1	NCM2_MOUSE	O35136	mus musculus
39	407	5.9	873	1	FS21_DROME	P34082	drosophila
40	398	5.8	1271	1	MYPC_CHICK	O90688	gallus gall
41	396	5.8	1913	1	KMLS_HUMAN	O15746	homo sapien
42	393	5.7	898	1	FAS2_SCHAM	P22648	schistocerc
43	392.5	5.7	2481	1	UN52_CAREL	Q06561	caenorhabdi
44	381.5	5.6	1906	1	KMLS_CHICK	P11799	gallus gall
45	370	5.4	1131	1	MYPC_CHICK	P16419	gallus gall

ALIGNMENTS

RESULT 1
NEOL_MOUSE
ID NEOL_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEOL OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.N.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis.";
RL Oocyte 15:691-700(1997).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4 AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UBQUITOUSLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E16.5.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR ENBL: Y0535; CAA70727.1; -
DR HSP: P02751; 117G.
DR MCD: MGI:1097159; NEOL.
DR INTERPRO: IPR001777; -
DR INTERPRO: IPR003006; -
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; sg; 4.

Page 3

[illegible]

Query Match 9.4%; Score 642; DB 1; Length 1377;
Best Local Similarity 22.0%; Pred. No. 9.4e-28;
Matches 325; Conservative 194; Mismatches 535; Indels 426; Gaps 64;

Qy 993 LGLYI-----NKSH-----VPIVIVAILLIIFVVIIVIAICWYN 926
 Db 1040 KGRLLPDLSGPKPKSPKSGNSPHGSPISPLDNLMLIVISVIGVITIVVIVIAVPCRR 1099
 Qy 927 SRNSGDKDRSFIKINDGSEVMASN-----NLW-----DVAQNPQNPQNTYAGR 970
 Db 1100 TTSQKKLRACKSVNGSHKYKGNCKDVKPDLNHHRELKPKDKSPQNPVMD--- 1156
 Qy 971 MTNNKNGQALYSITP-----NAQDFNNDYDSGTHRPHSGSHYHVAQLGQGNAMSTFY 1028
 Db 1157 -TPIPANSQ---DITPVDMSD-----SNHQRRNSYGHSE-----DSMTILA 1197
 Qy 1029 GNYTHDPSPTATTLVLNSQQAFLNDKMLRAPAMPTNPVPEPPARTADHAGRSRS 1088
 Db 1198 GRB-----CMRPM-----NPFDSQPPQSVNTPSTDTWASS 1232
 Qy 1089 SRASDGRGLNGLLHRTSGSRSDSPPTDVSIVQLHSSDGTGSSKTCGERTPPNKY 1148
 Db 1233 S-----QTCTDQDQPGATSSVYASSQSDSGSLPTAHVR-----PSHP 1274
 Qy 1149 LMDFIPPPSNPPPGGHVDTA-----TRQLARGSTPREDTYSVSDGAFADVWN 1201
 Db 1275 LKSFVFA-----IPFGPIITDPALESTPLLSQALNH-----RLHSVTSASITGLG--- 1323
 Qy 1202 ARPSTSRNULGGRPLAGKRRDSDSSSLMDDSDGSSSEAD 1241
 Db 1324 SRFP-----NPFVVPSPAPVQENTMLESDSESYFED 1355

RESULT 3

NEOL_HUMAN

ID NEOL_HUMAN STANDARD; PRT; 1461 AA.
 AC Q92859; C00340;
 DT 01-OCT-2000 (rel. 40, Created)
 DT 01-OCT-2000 (rel. 40, Last sequence update)
 DT 01-OCT-2000 (rel. 40, Last annotation update)
 DE NEOGENIN PRECURSOR.
 GN NEOL OR NGN.
 OS Homo sapiens; (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=PETAL;BRAIN;
 RX MEDLINE=97216553; Pubmed=9121761;
 RA Meyerhardt J.A., Look A.T., Blyner S.H., Pearson E.R.;
 RT "Identification and characterization of neogenin, a DCC-related
 RT gene.";
 RL Oncogene 14:1129-1136(1997).
 [2]
 SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=PETAL BRAIN;
 RX MEDLINE=97312699; Pubmed=9169140;
 RA Vielmetter J., Chen X.-N., Miskewich F., Lane R.P., Yamakawa K.,
 RA Korenberg J.R., Dreyer W.J.;
 RT "Molecular characterization of human neogenin, a DCC-related protein,
 RT and the mapping of its gene (NEOL) to chromosomal position 15q22.3-
 RT q23.";
 RL Genomics 41:414-421(1997).

CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
 CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
 CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
 CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
 CC LINES.
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
 CC TUMOR SUPPRESSOR PROTEIN DCC.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U61262; AB017263.1;
 DR EMBL: U72391; ACN51287.1;
 DR MIM: 601907;
 DR HSSP: Q02751; 1776;
 DR INTERPRO: IPR001777;
 DR INTERPRO: IPR003006;
 DR PFM: PP00041; fn3; 6;
 DR PFM: PP00047; fn4; 4;
 DR PRINTS: PR00014; FNTYPEII.
 KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 34 1461 NEOGENIN.
 FT DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1106 1126 POTENTIAL.
 FT DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 67 136 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 166 228 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 263 327 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 355 417 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 436 533 FIBRONECTIN TYPE-III.
 FT DOMAIN 536 629 FIBRONECTIN TYPE-III.
 FT DOMAIN 630 729 FIBRONECTIN TYPE-III.
 FT DOMAIN 735 829 FIBRONECTIN TYPE-III.
 FT DOMAIN 850 950 FIBRONECTIN TYPE-III.
 FT DOMAIN 951 1052 FIBRONECTIN TYPE-III.
 FT DOMAIN 1118 1121 POLY-VAR.
 FT DISULFID 74 129 BY SIMILARITY.
 FT DISULFID 173 221 BY SIMILARITY.
 FT DISULFID 270 320 BY SIMILARITY.
 FT DISULFID 362 410 BY SIMILARITY.
 FT CARBOHYD 73 73 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 210 210 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 326 326 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 715 715 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 909 909 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPIC 1248 1300 MISSING (IN ISOFORM 2).
 FT CONFLICT 168 168 G -> N (IN REF. 2).
 SQ SEQUENCE 1461 AA; 159958 MW; 7AAE897E69635A21 CRC64;

Query Match 9.3%; Score 639.5; DB 1; Length 1461;
 Best Local Similarity 21.6%; Pred. No. 1.4e-27;
 Matches 327; Conservative 202; Mismatches 546; Indels 437; Gaps 65;

Qy 36 PIDVYVSGSPATLNGCA--KPSTAKITVYKQDPVITNKQV-NSHRIVLDSGLFLK 92
 Db 59 PVDLTVSGVSGVILNCSAYSEPS-PKIDWKDQ---TFLMLVSDDRQLLPGSLFTSN 113
 Qy 93 VNSGKNGKSDGAGYCVASNEH-GEVKSNEGSLKMLKLEDFRVRPRTVQALGEMAVL 151
 Db 114 VVSKNKH-PUGETQCVATVESLGTISRTAKILVAGLPK-PTSQPEPSSVYAGNAIL 171
 Qy 152 ECSPGPFPEVYVSKRDKDELRIQDMPRYTLASDNLNIDVDSQSDQCVANNNG 211
 Db 172 NCEVNDL-VFVFWQNRQQLLLDD-RVILKPSGLMVISNATGDDGLKCVVSGSP 228
 Qy 212 ERVSNPRLSVTEPK-----FEQEPKMTVDVGAALDFCRVTDQDPQITWKRKNP 265
 Db 229 PKYSDEVKLVLPQPEVLSDLVFLQKPSPLVRYGQDVLFCVAGSLPTPTIKNNKEA 288
 Qy 266 MPV--TRATIKDNRGLRIERVQSDSEGYCVARNPNTLEASHLVQAPPSPYTKA 323

Db 289 LUTSESRNLVLAGGSLEISDVTEDDAGTYFCADMGNETPAQAEITVQAOPEFLKQPT 348
 Qy 324 DQSVAGGATTAPECTVLGQSPATFWSKEQDQLLPSTVSADGRTKVSPTGLTIEBVR 383
 Db 349 NITAEHSMIDVPECEYTGKPTVKKVWG--DWIPVS-----DTRFKVKEHLQVLGLV 401
 Qy 384 QVDEGATVCAQNSAGSSSLKAL-----KATFET 413
 Db 402 KSDEGPTQCIANDVGNAGQAQLLILERAPATTGLPSAPRDVVASLSTRFKILWRT 461
 Qy 414 KG-----RVKKKSKMGKQKKN-----VQSIKTLISAVT-----GN 446
 Db 462 PASDPGRDNLITVPTTKEGIARERVENTSHPGEMQVITQNLMPATVIFRVMQKNGHS 521
 Qy 447 TPAKPTPIIEHGHQWQTLVGSALPQASGKPTPGISN-----LRD 489
 Db 522 GESSAPLRE--TOPEVLQPGAPMLRAYAASPTISITVWETVPSGNGEIQNTLYNEX 579
 Qy 490 GLPIDITDSRISQBSTGSLHDLAKPDQTVYTICKAKNDEGSTWASLIVEDHTSNAQF 549
 Db 580 GTDKE-QDVVDSH--SYTINGLAKTYEYSPFVAVNKNKGQVSTPQVAVN----- 627
 Qy 550 VMDPDSNPSFPTPTIIVNYDTE-VELHN--APSTSGAGITGVIITQY----- 597
 Db 628 ----TISDVPFAAPQNLSEVANSKSIHMQPAPATON-QITQTKIKIRKASRSDV 682
 Qy 598 -----TSPDLQQTWFIIPD 611
 Db 683 TETLVSTQLSQLIEGLDRGTENFRVAALTINGTGPATWLSAETPSDLOET--RVPE 740
 Qy 612 TVASTE-----TRIKGLK 624
 Db 741 VPSLLHVRPLVTSIVSVTPPENQIVVGYALGYGIGSPHAQTIKVDYQRTYITELD 800
 Qy 625 PSHSTVIRANEKIGTQPSVSSALVTSKPAQVALSDKNKMDIAEKRLT-----S 679
 Db 801 PSHSTVITLKPNWGGEGIPLYESAV--TRH-----TDTSEVLEHNAPTVPVPT 852
 Qy 680 EQLIKL-EVKYTIINSTAVRLFKKRL-----EELDQ--EYIKWGGPPH-----DNQTVNV 730
 Db 853 PMHPPGVGVGSILSHDITRIWADNSLPKHKITDSRTYIVM--KINIPANTKYNA 908
 Qy 731 TSPSTENTVSNLMPFTIEFFVI--PYHSGVHSI--HGAPSNMDLIRAPASLPPE 785
 Db 909 NA-TILSTLVTLGLAPTLIEFSVYTKGRSSSTWMTAGT-----TFELVPTSPFN 959
 Qy 786 DWAI--RMALNLTIRISWAKPADGINGILKGPQIVVQAGPNNRNNITNERRASVTLF 843
 Db 960 DTVTSKSGKPKTILVWQPPSE--ANGKITG--IIYSTDVNAELDNWTEVPGVNGRLT 1016
 Qy 844 HLVTGMT--YKIRVAARSNGGVGSGTSEVIN-----NQTLEKHLAQQENESF 892
 Db 1017 WQIQLVLTPTTFYFKIQRNSKGM--PMSEAVQRTPKADSDKMDNDQASGSGGKG 1073
 Qy 893 LGLINKSHVP-----VIVVAILLIIPWIIIACTYGNRSN 929
 Db 1074 RLPDLSGDKPPMGSNSPHSGPTSPDLSNMLLIIIVSGVITIVVVIIVACTRTIS 1133
 Qy 930 SDCKDSFKINDGVSVMASN-----NLW-----DVAQNPNQNYTINAGTGM 973
 Db 1134 HQKKAACACVNGSHYKTKGSKDVKPPDLNIBHERLEKPKDKSPDPNFMID--TP 1189
 Qy 974 NNRNQALYSITP--NAQDFNNCDYSOTMHRPGESEHHTYQAOTGGPNAMSTFYNG 1031
 Db 1190 IPNWS--DITPWNDSM-----SNHQRRNSTYRGESE-----DSMTLAGRR 1231
 Qy 1032 YHDPDSPIATTVLVSNQPAKLDKMLRAPAMTNVPPPE-----PARTADHTAGRS 1086
 Db 1232 -----GMRPKM--MPFDSQPPQVISAIRPISLNFHHHF 1266
 Qy 1087 RSRASRAGRGTL-----NGGLHRTSGQSRSDSPHTDVSTVQ----- 1124

Db 1267 SSSLASPARSELVHSGSPWPIGTSMLSDRANSTSESVNTPSTDMPASSQTCOTDQO 1326
 Qy 1125 ----LASSDGTGSKERTGERPT-----PNTKLMDFIPPPSPPPGPGVYDTA----- 1172
 Db 1327 PEGATSSSLASSQEEDSQSLPTAHVRPSHLKSFVPA-----IPPGPPYTDALPSTP 1383
 Qy 1172 --TRQLMGSTPREDTIVSDVSAGFARVDVNARPTSNRNLKGLCKKROODSQRSLS 1229
 Db 1384 LLSQALAH--HHSVKTASITGLR-----SRPP-----MPVVPSAPVQETTR 1427
 Qy 1230 MMDDDGSSSED 1241
 Db 1428 MLESDSSSTED 1439

RESULT 4

NEOL_CHICK

ID NEOL_CHICK STANDARD; PRT; 1443 AA.

AC Q90610;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NMOGENIN (FRAGMENT)

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WHITE LEGHORN; TISSUE=EMBRYONIC BRAIN;

RM MEDLINE=95105243; PubMed=7806578;

RA Vleminckx J., Roman J.M., Dreyer N.J.;

RT 'Neogenin, an avian cell surface protein expressed during terminal

RT neuronal differentiation, is closely related to the human tumor

RT suppressor molecule deleted in colorectal cancer.';

RL J. Cell Biol. 137:2009-2020(1994).

CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE

CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR

CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION

CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.

CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.

CC -1- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS

CC AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.

CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO

CC TUMOR SUPPRESSOR PROTEIN DCC.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>CC or send an email to license@isb-sib.ch).

CC

DR EMBL: U07644; AAC59662.1; -.

DR HSSP: P80362; 107L.

DR INTERPRO: IPR001777; -.

DR INTERPRO: IPR003006; -.

DR PFM: PF00041; fn3; 6.

DR PFM: PF00047; lg; 4.

KW Transmembrane; Immunoglobulin domain; Glycoprotein.

FT NON_TER 1 1

FT DOMAIN <1 1090 EXTRACELLULAR (POTENTIAL).

FT TRANSEM 1091 1111 POTENTIAL.

FT DOMAIN 1112 1443 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 33 102 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 132 194 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 229 293 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 321 383 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 422 519 FIBRONECTIN TYPE-III.

FT DOMAIN 522 615 FIBRONECTIN TYPE-III.

RA Lyons G.E., Korenberg J.R.;

FT	DOMAIN	885	972	FIBRONECTIN TYPE-III.
FT	DOMAIN	984	1076	FIBRONECTIN TYPE-III.
FT	DOMAIN	1088	1177	FIBRONECTIN TYPE-III.
FT	DOMAIN	1189	1273	FIBRONECTIN TYPE-III.
FT	DOMAIN	1300	1366	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	1380	1463	FIBRONECTIN TYPE-III.
FT	DOMAIN	1477	1562	FIBRONECTIN TYPE-III.
FT	DISULFID	46	102	BY SIMILARITY.
FT	DISULFID	145	197	BY SIMILARITY.
FT	DISULFID	246	293	BY SIMILARITY.
FT	DISULFID	335	385	BY SIMILARITY.
FT	DISULFID	428	484	BY SIMILARITY.
FT	DISULFID	525	575	BY SIMILARITY.
FT	DISULFID	617	669	BY SIMILARITY.
FT	DISULFID	711	766	BY SIMILARITY.
FT	DISULFID	809	865	BY SIMILARITY.
FT	DISULFID	1307	1359	BY SIMILARITY.
FT	CARBOHYD	28	28	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1562	1571	NFALINDYGS > KEAAARKKFS (IN SHORT ISOFORM).
FT	VARSPLIC	1572	2012	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	1893	2012	HRPDGILHPPLVIMDRLNNGGTSRDSLSGQACLEPP
FT				SGRLGQPTVLPDIPMEASASSTSGDQSGVQKAVATLPQ
FT				SHZELGQAQMKSSQESLSDSGRLGNQNPYSSTTLV
FT				IQGVTSYICLHLWETFC (IN REF. 1).
SQ	SEQUENCE	2012 AA;	222259 MR;	0B33CF87B1A08334 CRC64;

Query Match 8.78; Score 598; DB 1; Length 1212;
Best Local Similarity 25.28; Pred. No. 4.3e-25;
Matches 230; Conservative 131; Mismatches 365; Indels 188; Gaps

Qy	16	YINFDKIPNASVLNPIVHEPI-----DVVSGSGPATLGNKAKPS-TAKITWYKDG	66
Db	386	FVWRKDL-SAGDQVYVLEDGTPAIIISAFKSEVSPAPVSLNMGVQPLPTITWLD	444
Qy	67	QPVTINKEQVNSRI-----VLDTGSLF-LKLVNKGNGKNDSDAGATCYVASNEHGEYKSE	122
Db	445	DPILAG-----GSHRSQMITSEGNVYSLNIS--SQVSDGVTYRCANNSAGV-----	493
Qy	123	GSILKIAMLEDPF-----VRP-RTVQALGSGMAYLCSPPFGPEPVSRRKDLKRIJD	177
Db	494	-----LYQARIYVKGASIRPMKNITATAGRDYTIYIC-VIGYPTYSIKWYNSLNPFNH	548
Qy	178	MPRTILHSDGNLIDPVDRS-DSGTQCYANNMGVERYSV-----PARLSVFEPK	226
Db	549	R-QVAFENNGYLLKLSQVQKEVDEGEYTC-NVLVQPOLSTQSQSVTVYVVPPTQFPEP	605
Qy	227	KFEQSPKMDTVYDGAALVDFC-RVTGDQPOITWRRKNEMPMPTRAYIAKDN-----	281
Db	606	RP-----SIGQVFTEPCQVDFGDTITTWQDKRPIPSGLJ-VTIDNIDTSSLR	655
Qy	282	IERVQSDSGEYCYCYARNPAGTLSEASHRLVQAPSPQTPADQSPVAGGTAFECTUJL	341
Db	656	ISNLSMNGNHTTCIARNSAAVYHQSGLIVLVPPKPVYQPDQDGIQKAVILANSAG	701

Query Match 8.7%; Score 598; DB 1; Length 2012;
Best Local Similarity 25.2%; Pred. No. 4.3e-25;
Matches 230; Conservative 131; Mismatches 365; Indels 188; Gaps 41.

QY	16	YINFDININSLAPVYIEHT-----DVVSRGSPATLANKGKPS-TAKITWKYGD	56
Db	386	PVFRKDL-SQDQVQVLEDGPKIISAFSEKVSVPASVSLCNKVGTPPLPTITWLD	444
QY	67	QPVTINKEQWISHERI-----VLDTSGLF-LLKNVSKNGKSDAGATYCVASHEGYSKNE	122
Db	445	DPLTGG-----GSHRIQMTISBGNVYINLIS-----SQVRDGGVICTANNSAGV-----	493
QY	123	GSILKAMLREDFP-----VRP-RTVQALGGMVALECSPPRGFPFPPVSHRKDKELRID	177
Db	494	-----LYQARIYVNGASIRPMKNITAGRDYTIICR-VIGTYPYSIRKYNLNPFFNH	548
QY	178	MPRTYLSHSGNLIDPVDS-DSGTQYCAVNMVGEYSV-----PARLSVFEPK	226
Db	549	R-QVAFENGKLLKSDQKEQVEDGYTC-NVYVQQLSTQSIVHTVVPVFPFPFP	605
QY	227	KEQEPKDMTVDGAAVLEDC-RVYGDSPQLTHKNKMPMTIATKDN-----RGLR	281
Db	606	IRP-----SIQRFNVPVSCDGLPHTHAKRQVPPSPGTPGSLG-VITDMIDITPST	653
QY	282	IRVSPDGSDEGYCYARNKAPGTLASLHTRKQVAPSPGTPKQDQVSPAGVATCTEFTUJGL	341
Db	656	ISNLSINMNGNITITARNBAEVAHOSGLIVPVPPKVPVOPDQDYGKAVIKLNSAG	715

QY 342 QPSPAYFW-SKEGQDLFFSYVSADGRKTVPTGLTIEVYQVDSGATVCGMAGS 399
 Db 716 PTPVITWKFSGKAGVQPOP-IALNRQTVLSNGLLKHVVEEDSGVTLCKVSDVG 773
 QY 400 SLSLKAALKATFETKYGQKSKMGKQKQKNSIKYLSAVTGMTPAKPPPIEIGH 459
 Db 774 ADVSKS-----MVLTKI-----PAMITST 793
 QY 460 QMOTLW-GSSALPQASGKPTPGISLWROGLIP-----ITDSISQSTGSLHIA 511
 Db 794 PWTTLATQGGKMSCTAGKXPIIVWKEKDRINPEMARIIVSTKEVGEVISTQL 853
 QY 512 DLAKPGQVGTICIAKNEDGESTWSASLTVEDTNSAQVPMPSNTPSSPTQPIIVMT 571
 Db 854 PTVDEDSGFSCHAINSNGEDGRIQLIQV-----PPDPEIEIKDKV 897
 QY 572 DTEVELNNAPSTSCAGPTGVIQIQTSPDLQGTWNI-----PDVASTETIRGLK 624
 Db 898 ARTITLWMT-GFDGNSPTGYDIE-CXKNSDSWDSQARTKDVSPQLNSAT-----IIDIH 951
 QY 625 PSBSTYKMTINENKGIQTPSVSSALVTTSPKPAQVSLDKMKMDMAIAKRLTSLQ 984
 Db 952 PSSYTSIRMYAKNR-----IGKSEPSNELITADERAPG-----PPQVH 984
 QY 685 LEEVKINSTAVLFWK-KRLEL-LIDGYIAWGPPTNDNQYVW-----TSPSTN 737
 Db 993 LE-----PISSQIRVWAKPKHLLNGIIRGYQYGT-EYSTGQVQWISVDSGDSV 1048
 QY 738 VYVSNLWPTMYEFVPIYHVSVHIGAPNSMDVLA-EAPSLPDDVIRKMLT 795
 Db 1049 YTLNKKFTQYGLVQACRA-----GTGPFSSQITTELDVPSPPENQAIATPE 1103
 QY 796 TLAISKAPADGILAGQIVIVGAPNWN-----RWITNERAAVTLFHLVGMTY 851
 Db 1104 SISISNLSKALNALINGIQFVITWANMLDGLGEIAKITTTQ-PSLELDGLEYNTY 1161
 QY 852 KIRVAASNGSGVY 865
 Db 1162 SIQVLAFTAGDGY 1175

RESULT 6
 CONT_CHK
 ID CONT_CHK STANDARD; PRT; 1010 AA.
 AC P14781; P10450;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CONTACTIN PRECURSOR (NEURAL CELL RECOGNITION MOLECULE F1).
 SS Gallus gallus (Chicken).
 CC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=90180453; PubMed=2627374;
 RA Bruemendorf T., Wolff J.M., Rainer F., Rathjens F.G.;
 RT "Neural cell recognition molecule F1: homology with fibronectin type
 RT III and immunoglobulin type C domains."
 RL Neuron 2:1351-1361(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN;
 EX MEDLINE=89008597; PubMed=3049624;
 RA Ranscht B., Dours M.T.;
 RT "Sequence of contactin, a 130-kD glycoprotein concentrated in areas
 RT of interneuronal contact, defines a new member of the immunoglobulin
 RT supergene family in the nervous system."
 RL J. Cell Biol. 107:1561-1573(1988).
 RN [3]
 RP GPI-ANCHOR.
 EX MEDLINE=89286606; PubMed=2735929;
 RA Wolff J.M., Bruemendorf T., Rathjens F.G.;

RT "Neural cell recognition molecule F1: membrane interaction by
 RT covalently attached phosphatidylinositol."
 RL Biochem. Biophys. Res. Commun. 161:931-938(1989).
 CC 1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
 CC SYSTEM DEVELOPMENT.
 CC 1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC 1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC 1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC 1- CAUTION: REP. 2 SEQUENCE DIFFERENCES FROM THAT SHOWN IN THE C-TERMINUS
 CC AND IS LONGER DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X14877; CAA33018.1; -
 DR EMBL: Y00813; CAA68753.1; ALT_FRAME.
 DR PIR: J00094; J00094.
 DR PIR: S01998; S01998.
 DR HSP: P20241; ICFB.
 DR INTERPRO: IPR001777; -
 DR INTERPRO: IPR003006; -
 DR PFM: PF00044; fn3; 4.
 DR PFM: PF00047; ig; 6.
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 ? CONTACTIN.
 FT PROPEP 7 1010 REMOVED IN MATURE FORM.
 FT DOMAIN 50 113 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 143 210 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 247 308 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 336 389 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 420 482 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 510 581 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 593 599 GLY/PRO-RICH.
 FT DOMAIN 600 701 FIBRONECTIN TYPE-III.
 FT DOMAIN 702 803 FIBRONECTIN TYPE-III.
 FT DOMAIN 804 900 FIBRONECTIN TYPE-III.
 FT DOMAIN 901 996 FIBRONECTIN TYPE-III.
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1010 AA; 112507 MW; 2E38F071AF423AE1 CRC64;

Query Match 8.7%; Score 597; DB 1; Length 1010;
 Best Local Similarity 23.6%; Pred. No. 1.8e-25;
 Matches 268; Conservative 140; Mismatches 364; Indels 364; Gaps 47;

QY 6 FYRTHTHRTY-----NFKIPNANSLNPIEHPIDVW-----SGSPATLNGAKP 55
 Db 3 FFISHLVTLCTPCFVADSTHFESEEN-KYGPVPEPQDITYPEESDQGVMSNCRAR- 60
 QY 56 STAKITWYKDGQVITNKEQVNSHRIVL-DTGSLELLKYNKNGKSDAGATYCVASN 113
 Db 61 -----AVPPTKYKWLKNWDLTKDRYSVNGVRLVSNPKSRDAGKYVCVSN 110
 QY 114 EHEGVSNMGSGLKML-----REDFFVPRVTVQALGGDAVLECSPPRGVPEPV-VSWR 167
 Db 111 IFTVRSSEATLSFGYLDPPPEREYKVRK-----GVGVALCEPPIYVDDLSYRWL 165
 QY 168 KDKKELRIQ-DMPTRIHSDGNLLDPDRSDSGTYQCVANNVGERVSNR-RLSVFEK 225

Db 166 LNEFPVFIADRRRFSVQINGNLYIANVSDKNGSCF-----VSSPSITKSVSK 217
 Qy 226 ----PKFQE---PKDMIVD-----VGAALFDCRVTDGPQITWKRKNEMPMPT- 269
 Db 218 FIFLIPQADRAVYPAIDIKVFKDYALLGQVTLCPALGNVPYELNWSKLEPMAPA 277
 Qy 270 -----RAYI-----AKD--- 276
 Db 278 EISMAGVILKIFNIQYDEGLYECAENYKGDKHQAVYVQSPKWEHINDTKIDGS 337
 Qy 277 -----NRG-LIRIEVPQSDGEYCYTARNPAGTLEASAH 310
 Db 338 DLFWPCVATGKPIPIIRWLNKGVSRKGEIRIQGLFEDAGMYQCIANAGHIIYANAL 397
 Qy 311 RVQA-PPSPQTKPADQSPVA--GGTATFECTLVGQSPAVFWSKGQDQLFSPYSVAG 367
 Db 398 KIVASPTFELNPMKKKILAAKGRVILECKRPAAPKPKFSWSK-GTELLWGS----- 450
 Qy 368 RTKVSPTGLTIEEVRVDEGAYVCAHNSAGSSKALKATPETHGVQKSKMKQK 427
 Db 451 RHIMWDSLEIITWIKLDEGRYTCFAENNRKANGVYGLMEATAT- 497
 Qy 428 KQKNVSIKILISAVTGMTPKPPPIEHHQNTLWMSGSAIPQASGKPTPGIS-- 485
 Db 498 -----ITLAPLVNVTVGERNATQCIASHDPTLIDFT 530
 Qy 486 WLKRGLPIDIT-----DSRISQHSHTSLHLDKPDGYTCTIAKNEDEGSTWSASLT 539
 Db 531 WSLNGFVIFDEKEHEHYERNVMIKNSGELLIKVQLRHGRGRTCTAQTIVNSSASDLV 590
 Qy 540 VEDHTSNAQFVMPMDPNFSPPTQPIVNVDTVELHWNAPTSAGPITGIIQYTS 599
 Db 591 -----VRGP-----PGPGIRIKIEIRIDTAVALTWSR-GTDMHSPISKTIQ-SK 633
 Qy 600 POLQINWFNIPTVASTE-----YRIKGLKSHSYMFVIRAEENKIGTSPSVSA 649
 Db 634 TFLSEKWD-----AKTEPSDIEGNNESARVIDLFWMEYEPRIATNITGTGSPMSQ 688
 Qy 650 LVITSKPAQVALSD-----KKNMDMAIEK-----RL 677
 Db 689 RIRTEGAPPVAFPSDVGSGGGSNRELTIWMLPSREHYHNGWYVAFKPPGKEHNRV 748
 Qy 678 T-----SEQLIKE-----EV 688
 Db 749 TVTNPEIGTVHKDSMPPSTQTVKVAFNKSGDGPLFLANVTSQAQDPTVEVDV 808
 Qy 689 KTNSTAVFLFKARKKLELDGIYIK-WRGPRNTNDQVFNWVTSPTENY--VSNLMP 745
 Db 809 KVLSSSEISVSN-HVYTESVETQIRFWAHNDEKAAAGRVV--SQSTSKLENLAP 864
 Qy 746 FTWTEFFVPIFHSQVSHSGNSMDVLZARAPPSLPEDVIRMLNLTILR-----I 799
 Db 865 MYRTHDVSFNS--AGTGPESRTIDITKAPPGQR---RI---ISKVRSGRSRII 914
 Qy 800 SWNAKPADGKINGILGQIVIVGQAPNNNRNNTTNEAASVTLFWGTMYTKIRVAAS 859
 Db 915 TWQNVKAMSNESAVBGKTVLKPQDGHGKLFSTGKHTLEVP--VPSDGETVVEYHRS 971
 Qy 860 NGGWG-----YSRGTSEVMNQDLEKHLAAQENESFLGLINKSHVPIVIVA 909
 Db 972 BGGDGEVAQIKISGATAG-----PTLLGLV-----LPALGVLA 1006

RESULT 7

CONT_MOUSE

ID CONT_MOUSE STANDARD; PRT; 1020 AA.

AC P12960;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CONTACTIN PRECURSOR (NEURAL CELL SURFACE PROTEIN F3).

GN CNTN1.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Soliurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=BRAIN;
 RX MEDLINE=89340657; PubMed=2474555;
 RA Gennarini G., Cibelli G., Rougon G., Mattel M.-G., Goridis C.;
 RT "The mouse neuronal cell surface protein F3: a phosphatidylinositol-
 anchored member of the immunoglobulin superfamily related to chicken
 RT contactin.";
 RL J. Cell Biol. 109:775-788(1989).
 CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
 CC SYSTEM DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC -1- MISCELLANEOUS: F3 SHARES WITH L1, N-CAM, MAG, AND OTHER CELL
 CC ADHESION MOLECULES FROM NERVOUS TISSUE THE L2/HNK-1 CARBOHYDRATE
 CC EPITOPE.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC -----
 DR EMBL: X14943; CAA33075.1; .
 DR PIR: S05944; S05944.
 DR HSP: P40189; 180U.
 DR MD: MG1:105980; CNTN1.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PTM: P00041; fn3; 4.
 DR PTM: P00047; lg; 6.
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1 20
 FT CHAIN 21 ? CONTACTIN.
 FT PROPEP 7 1020 REMOVED IN MATURE FORM.
 FT DOMAIN 58 121 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 151 218 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 317 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 345 398 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 429 491 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 519 592 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 604 611 GLY/PRO-RICH.
 FT DOMAIN 611 712 FIBRONECTIN TYPE-III.
 FT DOMAIN 713 814 FIBRONECTIN TYPE-III.
 FT DOMAIN 815 910 FIBRONECTIN TYPE-III.
 FT DOMAIN 911 1006 FIBRONECTIN TYPE-III.
 FT DISULFID 65: 114 BY SIMILARITY.
 FT DISULFID 158: 211 BY SIMILARITY.
 FT DISULFID 263: 310 BY SIMILARITY.
 FT DISULFID 352 391 BY SIMILARITY.
 FT DISULFID 436 484 BY SIMILARITY.
 FT DISULFID 526 585 BY SIMILARITY.
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 521 521 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1020 AA; 11338 MW; 9DCDCAA0EAAACBCT CRC64;

Query Match 8.7%; Score 595; DB 1; Length 1020;
 Best Local Similarity 24.0%; Pred. No. 2.3e-25;

[illegible]

FT	DOMAIN	813	908
----	--------	-----	-----

CC
DR EMBL: M31725; AAA42201.1; -.
DR PIR: A34695; A34695.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 4.
DR PFAM: PF00047; Ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion. Repeat.

```

FT SIGNAL      1    30
FT CHAIN       31  21015 AXONIN-1.
FT PROPEP      21016 1040 REMOVED IN MATURE FORM.
FT DOMAIN      56   120 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      150  218 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      256  315 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      343  404 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      435  497 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      525  596 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN      608  614 GLY/PRO-RICH.
FT DOMAIN      613  708 FIBRONECTIN TYPE-III.
FT DOMAIN      716  811 FIBRONECTIN TYPE-III.
FT DOMAIN      818  910 FIBRONECTIN TYPE-III.
FT DOMAIN      911 1005 FIBRONECTIN TYPE-III.
FT SITE        796  798 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD    78   78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    200  200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    206  206 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    463  463 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    479  479 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    500  500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    527  527 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    777  777 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    832  832 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    920  920 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD    942  942 N-LINKED (GLCNAC. .) (POTENTIAL).
9Q SEQUENCE    1040 AA; 113042 MW; 6E707E6614CB4FB CRC64;

```

```

Query Match      8.5%; Score 585; DB 1; Length 1040;
Best Local Similarity 24.0%; Pred. No. 8.6e-25;
Matches 238; Conservative 113; Mismatches 381; Indels 258; Gaps 36;

```

```

Qy 26 SNLAPVIEHPIDVWV---SRGSPATLNCAGKPS-TAKITWKKDQGVVINKQVNSHRI 81
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 35 ATFGPIEDQPIGLLFPPEESAEQVTLACARASPPATYKRWKNGTDM- NLEPGSRHQL 92

Qy 82 VLDGSLFLKVGKNGKSDAGAYTCVSNHEGVEVSNBGLSLKMLAE-----132
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 93 M-GGNVIMSPT-----KTQDAGVYCCLASNPVTGVSKAVLRFGLQFSEKDEPQV 145

Qy 133 -----132
Db 146 KYHGEKGVMLPCNPARTPGSLFYRWLLNEFPNPTPTDGRHFPVSTGTLNLIARTNASDLG 205

Qy 133 -----DFVR-----PRTVQALGEMAVLESCP 155
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 206 NYSCLATSEMDFTSVKSFKAQLMLAAREDLPLFAPISAKRPPTETALVGOVLECF- 264

156 PRGFPPEVSNKRDDELRIQDMFRITLHSDGNLIIDVDRSDGTYCVANNMNGEVS 215
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 265 AFGNPVPRINKRVGSL-----SPQWAT-AEPTLQIYVSFDEDTCECAENSKG-RDT 318

Qy 216 NPARLSVFEKPKFEQPKDMTVDVGAALVDCRVTGDPQITWKRKNEMPHVTRIAYK 275
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 319 VQGRITVQAPFEMKLVISDTADIGSLNRCGCAAGCKPRPMVWNRKEP-----LAS 371

Qy 276 DNR-----GLRIERVQPSDEGEYCYARNPAGTLBAASHLQVA-PFSPQTKPADQSV 328
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 372 QNFVEVLADGIFSKLSLEDSGMQCVAKNKHGTIVASLQVQALAPDFQNPVRRILP 431

Qy 329 A--GGTAFECTLVGQSPSPAYFWSKESQDQLLFPSYVSADGRTKVSPGTGLTIEBVRQD 386
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 432 AARGGEISILQFPAAPKATILWSKGT-----LIGNSTVTVISDGLTILNLSRSD 484

Qy 387 BGATVCAGMNSAGSSLSKALKATFETKRGVKKKSKMGKQKQKNGVQSIILYLSAVTGN 446
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 485 BGKTYCPAENMGKANSTGLSVDAET-----512

Qy 447 TPAKPPPTIEHGHQNTLMVSSAILPCOAGSKPGPI--SWLRDGLPIDTD-----SR 499
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 513 -----ITLAFSSADINVGDNLTLCASHSDPTMDLTFTWTLDFDPIFKRPGHYRR 564

```

```

Qy 500 ISQSHT-GSLHJADLKPKDPTVYTCIAKNEDGSTWASLVEDHTSNQVFRMPOPSNF 558
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 565 ASAKETIGDLTLAHAVRHGKATCYCA-----TVDGTSKEATVLRAGP---609

Qy 559 PSSPTQPIVNVTDTEVELHWNAPSTSGAGPIYIIQYYSFDLQTFWNPITVASTE- 617
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 PGPGGVVVRDQITVQLSWSR-GPDMNSPIAKYTLQARTPPSGK-WKQRTNPVNIIEG 667

Qy 618 ---YIKGLGPSSTSYFVTRANEKGIOTPSVSSALVTSKPAQVALSDKNKMDMAIA 673
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 668 NASTAQVLGLPMWDETFRVASANILCTGEPSGSSKIRTKKAVSPVAPSLSGGGGARG 727

Qy 674 EKRLTSQKLKLEBKINSTAVLRWKKRKLLELD--GYIKWRGPRTNOMQVNVVT 731
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 728 E-----INWTPVSRETQNG-----DGFYLLSFR-----RQSSSQWJAR 764

Qy 732 SPSTE-NYVV---SNLMPPTNYEFPIPHSGVSHIGAPSNMDVLTAEPNPLPEDW 787
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 765 VPGADQYPIYVNDISQPTTPPEVKIRSYN---RRQGPESLTVLSAEEEPVAPAKV 821

Qy 788 RIRMLATRLTSKAPKADGINKLKGQVIVGQAPNNNNITWERRASVTLFHLVT 847
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 822 WAGSSSSSENMVSWE-PVLQDNGILLGYEIRI---WKAGDNEAADNVRAGTLOSARTV 878

Qy 848 GNT---YIKVVAARSNGGVVSHSTSEVI 873
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 879 GLNPNTKTHVYATNRACTGPAFSPADAM 908

```

```

RESULT 10
CAML_HUMAN
ID CAML_HUMAN STANDARD; PRT; 1257 AA.
AC P32004;
UT 01-JUL-1993 (Rel. 26, Created)
UT 01-OCT-1996 (Rel. 34, Last sequence update)
UT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1 OR N1CS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92031698; PubMed=1932117;
RA Kobayashi M., Miura M., Asou R., Oyemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of
RT different origin.";
RL Biochim. Biophys. Acta 1090:238-240 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Rosenthal A., Coutelle O., Drescher B.;
RL Submitted (APR-1994) to the ENBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=9229299; PubMed=1627459;
RA Reid R.A., Hemperly J.J.;
RT "Variants of human L1 cell adhesion molecule arise through alternate
RT splicing of RNA.";
RL J. Mol. Neurosci. 3:127-135 (1992).
RN [4]
RP SEQUENCE OF 353-1176 FROM N.A.
RX MEDLINE=9220233; PubMed=1923824;
RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
RT "PCR walking from microdissection clone M54 identifies three exons
RT from the human gene for the neural cell adhesion molecule L1
RT (CAM-L1).";
RL Nucleic Acids Res. 19:5395-5401 (1991).
RN [5]
RP SEQUENCE OF 332-371 FROM N.A.
RX MEDLINE=90353957; PubMed=2387585;
RA Djabeli M., Mattei M.-G., Nguyen C., Roux D., Demangeot J.,
RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;

```

- RT "The gene encoding LI, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [6]
RP SEQUENCE OF 1030-1257 FROM N.A.
RX MEDLINE-91132183; PubMed-1993895;
RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
RA Stallcup W.B.;
RT "Isolation and sequence of partial cDNA clones of human LI: homology
RT of human and rodent LI in the cytoplasmic region.";
RL J. Neurochem. 56:797-804(1991).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE-88298876; PubMed-3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule LI.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP VARIANT HSAS TYR-264.
RX MEDLINE-94004956; PubMed-8401576;
RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnal D.;
RT "A missense mutation confirms the LI defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANT HSAS/MASA LEU-1194.
RX MEDLINE-95187172; PubMed-7881431;
RA Fransen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the LICAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE-95004608; PubMed-7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metzenberg A., Ionescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPGL), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the LI gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE-95004609; PubMed-7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
RA Reyniers E., Korn B., Poustka A., Wilson G., Schrander-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT LICAM.";
RL Nat. Genet. 7:408-413(1994).
RN [12]
RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
RX MEDLINE-95282776; PubMed-7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Wadellius C., Kenwick S.;
RT "New domains of neural cell-adhesion molecule LI implicated in
RT X-linked hydrocephalus and MASA syndrome.";
RL Am. J. Hum. Genet. 56:1304-1314(1995).
RN [13]
RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
RX MEDLINE-96153146; PubMed-8556302;
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, LI.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]
RP VARIANTS HSAS/MASA/SPGL SER-179 AND ARG-370.
RX MEDLINE-96057512; PubMed-7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in LI-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE-97083370; PubMed-8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the LICAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; V-439-T-443 DEL; C-784 AND L-936-L-948 DEL.
RX MEDLINE-97338664; PubMed-9195224;
RA Macfarlane J.R., Du J.-S., Pepps M.E., Ramsden S., Donnal D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel LI CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).
RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE-98180721; PubMed-9521424;
RA Du J.-S., Srivastava A.K., Schwartz C.B.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF) detection of six novel mutations in the LI cell
RT adhesion molecule (LICAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE-98112489; PubMed-9452210;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
RL Hum. Mutat. Suppl. 1:5284-5287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE-98415726; PubMed-9744477;
RA Saugier-Verber T., Martin C., le Neur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
RT "Identification of novel LICAM mutations using fluorescence-assisted
RT mismatch analysis.";
RL Hum. Mutat. 12:259-266(1998).
CC CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC CC -1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF THREE X-LINKED
CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
CC SYLVIIUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBES.
CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
CC Milder presentation and a longer life expectancy. 3: SPASTIC
CC PARAPLEGIA TYPE 1 (SPGL). COLLECTIVELY THESE SYNDROMES ARE ALSO
CC KNOWN AS CRASH SYNDROME, AN ACRONYM WHICH STANDS FOR CORPUS
CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBES,
CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
CC CC -1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF HIRSCHSPRUNG DISEASE
CC (HSCR).
CC CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC CC -1- DATABASE: NAME LICAM; NOTE LICAM mutation Web Page;
CC CC WWW "http://hgins.ula.ac.be/dnalab/11".
CC

PI SIGNAL 1 20

Qy	79	HRLVPTDGLSVLKYNKSGKDSDAGAYCYVASNEHGEYSWEGSLALMA-----EDFR	135
		: : : : : : : : : : : : : : : : : : :	
Db	183	HFVSQTGCLYIARTNA-----SDLNGTSCIATS-HMDFSTSVSKFQAQLAAADTR	235
Qy	136	V-----RRTPQALOGEMAVLECSPPGPPEPVVSWKMDOKELIRQMPTYLHISDG	187
		: : : : : : : : : : : : : : : : : : :	
Db	236	LPASIKARFPATZYLVQQVVLYECF-AFGMVPRVRINRWVGSL-----SPQWT-APF	289
Qy	188	NLLIDVPDRDSGTQCVANNMGVEGVSPARLSVEEKKPFEQPKIMVDGGAANVLDC	247
		: : : : : : : : : : : : : : : : : : :	
Db	290	TLIQTSVSEDEGTGYECEAKNS-RDTWGRIIVQAPEMLAIVSDTDGSLNHWCK	348
Qy	248	RVTGDQPQQTIVWKKNKEHPVTRYIAKDN-----GLAIRVQSDEGEVTCYARNPA	301
		: : : : : : : : : : : : : : : : : : :	
Db	349	AAGAFRFPVFWMLNCF-----LASQRVSVLAGLOAFSFLSLEDGSMGTCAENKH	401
Qy	302	GTLASAHLRWA-PSPSQTKPDQSVPA-CGGATFECTVLQVQSPAIFVSKSGEQDLL	358
		: : : : : : : : : : : : : : : : : : :	
Db	402	GTTASASLAQAALPDFRLNRVRLRIPAARGELILPCQAPRAAPKVAWSK-GTEILV	460
Qy	359	FPSYISDAGRTHKSPGTLITIEVRVQDGGAYCAGMSAGSSLSKAALKATFTGRVQ	418
		: : : : : : : : : : : : : : : : : : :	
Db	461	NSS-----RVYTIPODTLILRNISRSDEGKYTCFAFNHGKANSTGLISVRDATX	510
Qy	419	KKASKMKHQKQNVNISSIIKLIASVNTYPARPPTPIEHGHQMTLMVGSSAILPCQASG	478
		: : : : : : : : : : : : : : : : : : :	
Db	511	-----ITLAPSSADINGLMTLCQHASH	534
Qy	479	KPTPGT-SWLRDGLPIDTD-----SRISQST-GSLIHLDLKKPQTVYTICIAKEDGS	530
		: : : : : : : : : : : : : : : : : : :	
Db	535	DPTMDLTPTWLDOPIDFDKPGGHYRTINTVKETIGDITLINAQLRHGKATYCQAQ--	590
Qy	531	ESTWSASLVEDHTSNMQOVRMPDPSPNFSPSPQPIIVNYDTVEZHEHNAPSTSGAGPI	590
		: : : : : : : : : : : : : : : : : : :	
Db	591	-----TWDSASKEATVYRCP---PGPGGVVVVDIGDTTGLQNSR-GFNDSPSI	638
Qy	591	TGVIQYVSPDLQGTWFNPIDTYASTE-----YRIKGKLGHSNMVIRAENEKIGTFPS	645
		: : : : : : : : : : : : : : : : : : :	
Db	639	AKTYLQARTPAGK-WKQVRTNPNANIGEMAEYAQLGTLTPTMDTFFERVASINSLGTGPS	697
Qy	646	VSSALVTSKPAAQVSKDKNMMDATAEKRLTSELKILLEVKTINSTAVRLKWKRL	705

RESULT 12
 CAML_MOUSE
 ID CAML_MOUSE STANDARD; PRT; 1260 AA.
 AC P16827;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GS L1CAM OR CAM1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-BRAIN;
 RA MEDLINE=88318924; PubMed=3412448;
 RA Mos M., Tacke R., Scherer H., Teplow D., Fruh K., Schachner M.;
 RT "Neural adhesion molecule L1 as a member of the immunoglobulin
 RT superfamily with binding domains similar to fibronectin.";
 RL Nature 334:701-703 (1988).
 CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BOUND
 CC TO AXONIN ON NEURONS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collabora-
 CC tion between the Swiss Institute of Bioinformatics and the EMBL outstat-
 CC e at the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in for com-
 CC mercial use and this statement is not removed. Usage by and for com-
 CC mercial entities requires a license agreement (see <http://www.isb-sib.ch/anno>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X12875; CA031368.1; -.
 DR PIR: S05479; S05479.
 DR HSP: P20241; JCFB.
 DR MGD: MGI:96721; L1CAM.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFM: PFM00041; fn3; 4.
 DR PFM: PFM00047; ig; 6.
 DR PRINTS: PR000014; FNTPEIII.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; signal; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
 FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1124 1146 POTENTIAL.
 FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 346 410 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 440 503 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 531 599 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 827 896 FIBRONECTIN TYPE-III.
 FT DOMAIN 932 994 FIBRONECTIN TYPE-III.
 FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III.
 FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1180 1183 MISSING (IN SHORT ISOFORM)
 (BY SIMILARITY).
 SQ SEQUENCE 1260 AA; 140968 MW; 22857001CBA2538 CRC64;

Query Match 8.3%; Score 567.5; DB 1; Length 1260;
 Best Local Similarity 21.8%; Pred. No. 1e-23;
 Matches 271; Conservative 173; Mismatches 470; Indels 327; Gaps 51;

Qy 30 PVIIIE-PIDVVSRSGLATLCAK-PSTAKITWKD---OPVITKQVNSHRIYLD 84
 Db 35 FVITQSGPRLVFPDGLSLACARGPQBFNWKDTHKFP---KEELG---VVVH 87
 Qy 85 ---TGSLLKLVNKGKNDSDA---GAYTCVSAHEGKVSNGSLKAMLRDPHV 136
 Db 88 EAPYSQSTFI---EGNSFQAFQGIYCTASNKLTGMSHE---QLVABGAPK 136
 Qy 137 RP---RTVQALGEMVLESCPPAGFPFVVSNNKDKELAIQDMPTLHSDGNLID 192
 Db 137 WKETVKPVEYEGESVFLPCNPPSAAPRIYV-MNSKFDIDQDEVSMGONGOLTPA 195
 Qy 193 PVDKRS-----GTQCVANMVGERSVNARLSVPEKP-FPEEPKMTITVQV 239
 Db 196 NVLTSDNHSIDICNAHPTGTITIQEPIDLR-KPTNSMIDKPRILFTNNSSLVAL 254
 Qy 240 -GAAVLDFCRVQDQPPQITNKRKNEMPTVTRAYIANDNGLRIERVQPSDEGYTCYAR 298
 Db 255 QGSLILCBTAGFPPTIKWLPSDMPITDRVITQNHNTLQLLVNEDDGETCLAE 314
 Qy 299 NPAGTLEASHLRVQAPSPQTPADQPSVAGGTAFECTLVGQSPAYPWSKEGQDGL 358
 Db 315 NSLGSARHAYITVYAAPYWLQKQSHLYGCHETARLDQVGRPOEITWRNG---M 370
 Qy 359 PPSVTSADGRTVSPGTGLTIEEVKQDEGATYVACMNSAGSSLSKALKATPETHKRVQ 418
 Db 371 SNEYTKMDQYTRIB-QGSLILSNVQPTDMVTQCEARNQGLLLANAYITV-QLPARIL 428
 Qy 419 KKKSKMGKQKQNVQSTIKYLSAVTGMPPAKPPTIEHGHQNTLMV---GSSALQCA 476
 Db 429 TK-----DNQTMVSGSTATLLCKA 449
 Qy 477 SKGPTGSLVMDGLPDIITDSRISQHSSTGLHIALKPDGTGYTCAKNEDEGSTWSA 536
 Db 450 FGAPVPSVQWLDEBGTIVLQDERFFPYANGTSLIRDLQANDGRYFCQANDQNNWTILA 509

Qy 537 SLIVEDHTSNAPVWMP-----DPS----- 556
 Db 510 NLQVKEATITQGPSAIEKKGARVTTTCQASPPSLQASITWAGDGRDLQERGDSKTF 569
 Qy 557 -----NFP-----SSPTPIIVN-----VIDTE 574
 Db 570 IEDGKLVIGSLDYSDQGNVSCVSTELDEVSRAQLLVGSGPVPHELSDRHLAQSQ 629
 Qy 575 VELHMNAPSTSGAGPITVLIQYISPDIG-QTWN---LPDVASTETRIKGLPSHSTY 630
 Db 630 VHLSN-SPAREDSPIEXYDIEFEDKEMAPEKPTSLGKVPQNTSTTLK---LSPVTWT 685
 Qy 631 FVIRANEKGIOTPS-VSSALVTTSKPAQVALSDKNKDMAIEKRLTSEDLIKLEEV 689
 Db 686 FRTVATNKYQGPSPVSESVTPE---AAPEKNVDVR-GBGMETNNWVITKPLR 738
 Qy 690 TINSTAVRLVKKRKLLELDGYIYNKRGPPRTNDQNVVNTSPSTENYVSNLMPFTNY 749
 Db 739 WMD-----VNAPOIG---YRVQVQKQKQETWKRTVSDSPF---LVVSNWSTFVPY 783
 Qy 750 EFPVTPSVHSYTHGAPSNQSV-LTAAEPSPLPEDVIRMLNLTILKISWAPKADG 808
 Db 784 EIVV---QAVNMQKGFEPQVITIGSDEDTQVSELEDITIFNNSITLVNRRVQLAQ 839
 Qy 809 INGLIGAFQIV---VQAPANNMNITNE---RAASVTLFHLVGMTYTKIRVAARSN 860
 Db 840 VGLHAGITVHTWNGSQKRSKRIHSHIVPANTTSAILSGRLPSTSYHVEQAFNG 899
 Qy 861 GGVGVSE-----GTSEVMMDQTEHLKLAQKESFLYGLINKSHVPIVIVA 909
 Db 900 GLGPAESKFTSTPBGVGPHEAL-----HELCQSDULLHQPPLSHNGVLT--- 948
 Qy 910 ILILFVIIIACTVNRNSRSDGKDRPSKINDGVSVMASNNLMDVQNPQNPMTYAG 969
 Db 949 -----GILLSTHPVGESEKGLPFLNSD--PELATHL-----TNLAP----- 984
 Qy 970 RMTNMRNNGALYSLTNMQAFNCCDSYGTMRHPGSEHHYHQAQLTGSGGNA----- 1023
 Db 985 -----DLQTRFQLQATQO-----GGGPAIVREGG 1010
 Qy 1024 -MSTF---YGMQYHDDPSYATITLVLSNQPAWLNDMLRAPMPTNVFP--EPAP 1076
 Db 1011 TMLALPKPQPGNISATAGENTSIVSVWPKRGCNFRHLFK--ALPBGKVSPOHPOPO 1068
 Qy 1077 YADHTAGRRSSRSRSDGRLN---GGLHL---RTSGS 1109
 Db 1069 TVSTNGSSTQWNLQPDITKYLHLIKERVLHLHLDVNTGT 1109

RESULT 13
 AX01.CHICK
 ID AX01.CHICK STANDARD; PRT; 1036 AA.
 AC P28685;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE AXONIN-1 PRECURSOR.
 OS Gallus gallus (Chicken).
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC TISSUE=BRAIN;
 RX MEDLINE=92174898; PubMed=1311675;
 RA Zuelig R.A., Rader C., Schroeder A., Kalousek M.B.,
 von Bohlen Und Halbach F., Osterwalder T., Inan C., Stoeckli E.T.,
 RA Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
 RT "The axonally secreted cell adhesion molecule, axonin-1. Primary
 RT structure, immunoglobulin-like and fibronectin-type-III-like domains
 RT and glycosyl-phosphatidylinositol anchorage";
 RL Eur. J. Biochem. 204:453-463(1992).

QY	30	PVIIIEHDPVSVSRGS---PATLNCAGK---STAKITVWKDGPVITNKQVSHRVL	83
Db	32	PVPFEGPQHTLLPFGSAEKKVITCLCRANPPTATR--WKNN---TELAGGPDSDRYL	85
QY	84	DTGSLFLKLVKSNKGKSDAGAYVCSASNEGKSNESGLSAMLNR---	132
Db	86	VAGDLYI---SNPKAKDAGSYQCVATNARGTVVSRASLSPGLQFSAEDRPVKI	140
QY	133	-----	132
Db	141	TEGNGVMFTCSPPPHYPALSYKLLNLFNPFADGGRFVSQTGNLYIAKTEASDLGY	200
QY	133	-----DFRVR-----PRTVQALGEMAVLECSPPR	157
Db	201	SCPATSHIDITKSVSKFSGLSAAEDRAYSPSIKAPADYVALTGOMVTECP-A	258

[illegible]

```

NRCL14
AC NRCL14.CHICK STANDARD; PRT; 1284 AA.
ID P353j1;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NC-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RC STRAIN-WHITE LEISHMAN; TISSUE=EMBRYONIC BRAIN;
EX MEDLINE=91258407; PubMed=2045418;
RA Grumet M., Mauro V., Burgom M.P., Edelman G.M., Cunningham B.A.;
RT "Structure of a new nervous system glycoprotein, NR-CAM, and its
RT relationship to subunits of neural cell adhesion molecules."

```

```

QY 30 PVIIEH-PIDVVYSRGSPATLWCAK-----PSTAKITVYKDGQPVITNKQVNSHRLVD 84
Db 41 PTTQQSKPDVIVDPRENIVICQCAKGGPPPS---FSWTRNGT-----HFDID 85
QY 85 TGSLELLVNSNG-----KNGKDDA-GAIYCVASNEHGVEYNSBGLSA---LAMLRED 134
Db 86 KDAQVITKPNNGSLVWNVKNNGVAAEYGVYQCTARNRNGAASINWIVTRSPMLTKE 145
QY 135 RVRPRTVALGEGMAVLCSPPRGFPBPVYVSKKDKLEAIQDMPTLLHSDGNLIDPV 194
Db 146 KLEPHVRE--GDSLVLNCRPVPGLPPIITV-MONAPQLQPSRVSGNGLGLDIFYSV 202
QY 195 DRSDST-YQVCA--NMWGERVSNPARLSVE-KPKFQEPKDMT-----VDV-GA 241
Db 203 QPBEDRVYICIAREWHQITQQQRIQSVKVSFTKPTVERPPIVLLPMGSSNWSNLEKRN 262
QY 242 AVLFDRCVITQDPOPIWTKRKNEMPNPTATIAIKDNGLRIARVQPSDEGETVCIYARN 301
Db 263 VLLCEIAGIPLTPVVRNKEGEGELPANFTFENFKETLKLIDVSEADSGNYTKCANRTL 322
QY 301 GTEASAHRLQVAPSPQTKPADQSPAGCTATFECTLVQQSPAYFWSKSGQDLFLPS 361
Db 322 GSTHIVISVTKAAPVTITAFRNLVLSGEDDITLCIRANGKPNPSISWLNG-----VPI 377
QY 362 TVSADGRKTVSPPTGTLIEZRVQDDEAYCAVGMNSAGGSLSAALAKATFETKRVQYKK 421
Db 378 AIAPDESRKVDGDTIIFSVAQERSSAVYQCSNAWEG-----HFDID 435
QY 421 SKMGKQKQNVQSIKYLISAVTGNTPAKPPPTIEHGQWQITLVGSSAILPQCAQSGKPT 481
Db 435 -----YLANAFVNWIAEPPLILPANKLIVQVIADSPALIDCAVTFSPK 459
QY 481 PGISWLROGLRIDTDSISQNSHSGSLHIALKKPQTVGTCIAKNEHGESTWSASLTVE 541
Db 460 PELEFHWGVKSGSLRNGEYVFDHNGLEIPLVAQNSDGTTCYVARKNLAKTQNEVQLVE 519
QY 542 DHT-----SNAQFV-----RMPDPSNFPSSPTQPLI 567
Db 520 DPTMIKQFQVTKQRSQAQSPFVCIKHDPITLPTVWLADNNELPDERPLWKGNDKLI 579
QY 567 VNWITDE----- 579
Db 580 MNTTDDGDTTICVNTLSDVSAASVLVWVAAPPTAIYARNWPPDLDELZLTQLERS 639
QY 639 ELHNAPSTSGAGPIYGLIYQ-----YSPDLQWNTNIPVYASTEYIKELKPSHST 699
Db 640 ELSW-VPEENENSPVINNFVIEYEDGLHEPGWVHYQTEVQ--SHTTWQLK-LVSNWYPS 699
QY 699 VIRAENEGITGTVSSALVYITSPKAQVALLSKDNKMDIAAEKRLTSD-----LIKLEVE 688
Db 696 RVTAENEGISRPD-----SPSRYCLTKAMPENPNSWOGIGESPNONVITNESL 746

```

Qy 689 KTIINAVLFWKKRLLELDIGYIKWGPPTNDQNVYVTSPTSTVVSNNPMTN 748
 Db 747 KGFQSGSGPGLQYK-----VSWR--QKQVDEWTSVVVANVSKIYSGTPTFP 792

Qy 749 YEYFPIYHSGVHSIHGAPSNMDVL--TADAPSLPPEVDRIKMLKLTIRISNAPKA 806
 Db 793 YEYK--QALNDLQYAPES-EVIGHSGEDLPMWAPGVNVDVINSTIAKGVNDPVL 847

Qy 807 DGINGLKQGVIVGQAPNNRNITTHERAASV----FHLVTGN---TKYIRVA 857
 Db 848 KSVYSGELQYK-VYTKVQSLSRSSKHVEKLTIFRGNTKGLPGLEPSSYKLVNRV 906

Qy 858 RSGVGG 864
 Db 907 VNGGEG 913

RESULT 15
 CAML_RAT
 CAML_RAT STANDARD: PRT; 1259 AA.
 AC Q05695;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN LICAM OR CAML1.
 OS Rattus norvegicus (Rat).
 CC Pukyryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:91372414; PubMed:1894011;
 RA Miura M., Kobayashi M., Asou H., Oyemura K.;
 RT 'Molecular cloning of cDNA encoding the rat neural cell adhesion
 RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced
 RT by differential splicing';
 RL FEBS Lett. 289:91-95(1991).

CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING.
 CC -1- TISSUE SPECIFICITY: THE SHORTER ISOFORM IS PREDOMINANTLY FOUND IN
 CC THE BRAIN, WHILE THE LONGER ISOFORM IS FOUND IN THE PERIPHERAL
 CC NERVOUS SYSTEM.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X59149; CAM1860.1; .
 DR PIR: S17655; S17655.
 DR HSP: P20241; 1CFB.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 4.
 DR PFAM: PF00047; ig; 6.
 DR PRINTS: PR00014; FNTPEI11.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 1259 NEURAL CELL ADHESION MOLECULE L1.
 FT DOMAIN 20 1122 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1123 1145 POTENTIAL.
 FT DOMAIN 1146 1259 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 346 410 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 440 503 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 531 599 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 827 896 FIBRONECTIN TYPE-III.
 FT DOMAIN 932 994 FIBRONECTIN TYPE-III.
 FT DOMAIN 1032 1093 FIBRONECTIN TYPE-III.
 FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 824 824 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 875 875 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 968 968 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 978 978 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1021 1021 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPIC 1179 1182 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 1259 AA; 140934 MW; 196818022D8F24AB CRC64;

Query Match 8.18; Score 557; DB 1: Length 1259;
 Best Local Similarity 22.38; Pred. No. 3.9e-23;
 Matches 244; Conservative 156; Mismatches 429; Indels 264; Gaps 42;

Qy 30 PVIIIE-PIDVWVSRGSPATLACGAK-PSAKTIWKDG---QPVITNKQVNSHRIYLD 84
 Db 35 PVITGSPRHLVPTDITSLACAGRGQVYFWKXDGIFKP---KEELG---VYVH 87

Qy 85 ---TGSLLFVWNGKNGKSDA---GAYTCVSNWGEVKSNGEGLKAMLEDFRV 136
 Db 88 EAPITSGSFTI-----EGNNSPQRFQITRCATSNNGTAMSH---IQLVAGAPK 136

Qy 137 RP---RYVQALGGMVLECSPPGFPPEVVSWRADKELRIQDPYTLASGNIID 192
 Db 137 WKETVYKPEVEGESEVLPCKPPPSAALRIYV-WNSKLHILQDERVSMGNDLFLA 195

Qy 193 PVDKRS-----GTQCVANNMVERSNPARLSVFEPKPF---EQEPMVDV 238
 Db 196 NVLTSQNSQSDICNHPFGPTITIQEPIDLR-KPTNSMIDKPRLLPPTNSSSHVAL 254

Qy 239 VGAALFDCRVTGDPQPTIKRWKNPMTVTRATYAKDNRLRIERVQPSDEGYCTYCR 298
 Db 255 QGSLILCEIAGFPPTIKWLPSPDPMTDRVYQNHNTLQLLVGEEDDGYETCLAE 314

Qy 299 NPAGTLEASALRWQAPSPQTKPADQSPVAGTATFECTLVGQSPATFWSKEQQQL 358
 Db 315 NLSGSSARAYTVYAAPYWLQPGSHLYGPGETARLDQVQGRPQPEYTWKING---M 370

Qy 359 PFSYVSADGRKIVSGPTGLTIEEVQVQDEGAYCAGNAGSSLSKALKATFETKRGV 418
 Db 371 SIEKYWKQKTRIE-QGSLILSNVQSPDTMTQCEARNQGHLLANVITVY-QIPARIL 428

Qy 419 KKSCKMKGKQNNVQSLIKLISAVTGNPAKPFPIIEHQHQNQTMV--GSSALLPCA 476
 Db 428 TK-----DNQTYAVGSGTATLLCKA 449

Qy 477 SGKPTGISLWDLGPDIDITSRISQSGSLHIALDKPQGTIVYCIANNDEGSTWSA 536
 Db 450 FGAPVPSVWLEDEGTTVLDQERFFPYANGHLIROLQNDGTGTFQANDQNWITL 569

Search completed: January 22, 2001, 12:28:41
Job time: 1222 sec

Qy 84 DTGSLFLLKVN⁸⁴SGKNGKDS⁸⁵DAGAY⁸⁶YCVASNEHGEVKS⁸⁷NEGSLK⁸⁸LAMLRED⁸⁹FRVR⁹⁰PRT⁹¹VQA 143

Qy 144 LGGEMAVLCSPGPGFPPEPVYWRKDDKELIQDMPTRYTLHSDGMLIDPVDASDSGTQ 203
 Db 145 LGGEMAVLCSPGPGFPPEPVYWRKDDKELIQDMPTRYTLHSDGMLIDPVDASDSGTQ 204

Qy 204 CVANNMYGVRNSPARLSVEPKFPEQPKDPTVGVGAALVDCRVGTGDPQITWKRN 263
 Db 205 CVANNMYGVRNSPARLSVEPKFPEQPKDPTVGVGAALVDCRVGTGDPQITWKRN 264

Qy 264 EPMPTVTRAYIAKNRGLRIERVQSDSEGTVCYARNAGTLEASAHVQAPPSQTKPA 323
 Db 265 EPMPTVTRAYIAKNRGLRIERVQSDSEGTVCYARNAGTLEASAHVQAPPSQTKPA 324

Qy 324 DQSVAGAGTACTEFTLGVQSPATYFWSKSGQDQLLPSTVSADGRTKVSPTGLTIEVR 383
 Db 325 DQSVAGAGTACTEFTLGVQSPATYFWSKSGQDQLLPSTVSADGRTKVSPTGLTIEVR 384

Qy 384 QVDEGATVCAHNSAGSSLSKALKATPFTKGRVQKSKSKGQKQKQVQSIKYLISAV 443
 Db 385 QVDEGATVCAHNSAGSSLSKALKATV-TT-----AV 415

Qy 444 TGNTPAKPPTTIEHGNQMTLWVSSAILPQASGKPTGSLWLDGLPIDITDSRISGN 503
 Db 416 TGNTPAKPPTTIEHGNQMTLWVSSAILPQASGKPTGSLWLDGLPIDITDSRISGN 475

Qy 504 STGSLIADLKKPDTGVYTCIAKNEGESTWSALVDEHNSAQVVMPPDSNFPSSPT 563
 Db 476 STGSLIADLKKPDTGVYTCIAKNEGESTWSALVDEHNSAQVVMPPDSNFPSSPT 535

Qy 564 QPITVWNTDEVELHWNAPSTSGAGPTGYLIQYSPDLGQTFWNPIDVASTERYIKGL 623
 Db 536 QPITVWNTDEVELHWNAPSTSGAGPTGYLIQYSPDLGQTFWNPIDVASTERYIKGL 595

Qy 624 KPSHSVMFVIRAKNEKIGTPSVSSALVTSKPAQVALSDANKNDIAEKALTSQILI 683
 Db 596 KPSHSVMFVIRAKNEKIGTPSVSSALVTSKPAQVALSDANKNDIAEKALTSQILI 655

Qy 684 KLEKVTINSTAVLFWKARKLEELIDGTYIKWGGPPTNDQNTVWVSTSTENTVSNL 743
 Db 656 KLEKVTINSTAVLFWKARKLEELIDGTYIKWGGPPTNDQNTVWVSTSTENTVSNL 715

Qy 744 MPTTNEFFVPIPSGVSNSHGAPSNMDVLTAEAPPSPPEVDVIRKMLTLTRISWKA 803
 Db 716 MPTTNEFFVPIPSGVSNSHGAPSNMDVLTAEAPPSPPEVDVIRKMLTLTRISWKA 775

Qy 804 PKADGILNGLAGFQIVIVGQAPNNNRNITNERAASVTLFHLVTGNTKIRVAASNGVG 863
 Db 776 PKADGILNGLAGFQIVIVGQAPNNNRNITNERAASVTLFHLVTGNTKIRVAASNGVG 835

Qy 864 GVSHTSEVMNQDTELEKHLAQBNESFLGLINKSHVPIVIVALLIFVVIITANCY 923
 Db 836 GVSHTSEVMNQDTELEKHLAQBNESFLGLINKSHVPIVIVALLIFVVIITANCY 895

Qy 924 WNSKNSDGRDPSFIKINDGVSMAKNLWDVAQNPQNMPTNAGMTNNKRNQALTS 983
 Db 896 WNSKNSDGRDPSFIKINDGVSMAKNLWDVAQNPQNMPTNAGMTNNKRNQALTS 955

Qy 984 LTPNAQDFNNCDQSTGMRHSGSEHXYAQLTGCGPNAMSTFTGNQYRDDPSPTATT 1043
 Db 956 LTPNAQDFNNCDQSTGMRHSGSEHXYAQLTGCGPNAMSTFTGNQYRDDPSPTATT 1015

Qy 1044 LVLSNQAPWLANDKLRAPAMPTNPVPPPPARYADTAGRSSRSRSDGRLNGLGH 1103
 Db 1016 LVLSNQAPWLANDKLRAPAMPTNPVPPPPARYADTAGRSSRSRSDGRLNGLGH 1075

Qy 1104 HRTSGQSQRSDPHTDVSYQLHSDGSGSKERTGERTPKNTLMDFIPPPSNWPPP 1163
 Db 1076 HRTSGQSQRSDPHTDVSYQLHSDGSGSKERTGERTPKNTLMDFIPPPSNWPPP 1135

Qy 1164 GSHVID----TATRRQLNKGSTPREDTYDSVSGAFARVDNARPTRSNRLNGRPLK 1219
 Db 1136 GSHVIDIDFQTATRRQLNKGSTPREDTYDSVSGAFARVDNARPTRSNRLNGRPLK 1195

Qy 1220 RDDSGSSSLMDGGSSSEADGSEGDVPGGVKAVPMSIASLHSCYTNGTA 1279

Db 1196 RDDSGSSSLMDGGSSSEADGSEGDVPGGVKAVPMSIASLHSCYTNGTA 1255

Qy 1280 QRFPSIRPNNGIVTQRT 1297

Db 1256 QRFPSIRPNNGIVTQRT 1273

RESULT 2
 C01632
 ID C01632 PRELIMINARY; PRT; 874 AA.
 AC C01632;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE CODED FOR BY C. ELEGANS CDNA CRESC12R.
 GN ZK377.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 EX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,
 RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roop A., Saunders D., Showkeen R.,
 RA Shuldron M., Smith A., Sonnenberg E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproot J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 RW [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Nhan M., Hawkins J.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
 RW [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
 RW [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: U88183; AAB52657.1; -.
 DR HSSP: P56276; IZLK.
 DR INTERPRO: IPRO01777; -.
 DR INTERPRO: IPRO03006; -.
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; ig; 1.
 DR PRINTS: PR00014; FNTYPEII.
 SQ SEQUENCE 874 AA; 95861 MW; BC72270818D734C9 CRC64;

Query Match 67.5%; Score 4628; DB 5; Length 874;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 MGKQKNQVSIKYLISAVTGNTPAKPPTTIEHGNQMTLWVSSAILPQASGKPTG 483

Db 1 MGKQKNQVSIKYLISAVTGNTPAKPPTTIEHGNQMTLWVSSAILPQASGKPTG 60

Qy 484 ISWLRLDGLPIDITDSRISGHTSGSLHIDLKLPDTGVYTCIAKNEGESTWSALVDEH 543

Db 61 ISMLRDGLPIDITDSRISQHSGLIHLADLKRPDTGYTTCIANKNEGSTWSALSTVDEH 120
Qy 544 TSMAGQVMPDPSPFPSSPTQPIIIVNVDTDEVLHNPASTSGAGPIGTIIQYITSPDLG 603
Db 121 TSMAGQVMPDPSPFPSSPTQPIIIVNVDTDEVLHNPASTSGAGPIGTIIQYITSPDLG 180
Qy 604 QTWNIPOTVASTETRIKGLKPSHSMYFVIRABNEKIGITGPSVSAVLTSPKPAQVALS 663
Db 181 QTWNIPOTVASTETRIKGLKPSHSMYFVIRABNEKIGITGPSVSAVLTSPKPAQVALS 240
Qy 664 DKMKDMAIAEKRLTSQLKILEEYKINSTAVRLFWKKRLEELIDGTYIKNRGPPPTN 723
Db 241 DKMKDMAIAEKRLTSQLKILEEYKINSTAVRLFWKKRLEELIDGTYIKNRGPPPTN 300
Qy 724 DNQIVNVTPSTENTYVNSMLPFTWYEFFVPIHSGVSHIGAPSNMDVLTAPPSLP 783
Db 301 DNQIVNVTPSTENTYVNSMLPFTWYEFFVPIHSGVSHIGAPSNMDVLTAPPSLP 360
784 PEDVIRMLNLTIRISWAKPAGDINGILKGPQIVVQAPNNRMTTNERAASVTLF 843
Db 361 PEDVIRMLNLTIRISWAKPAGDINGILKGPQIVVQAPNNRMTTNERAASVTLF 420
Qy 844 HLVGTMTKIRVAARSNGGVSHGTSSEVINQDTLEKLAQGNESFLYGLINKSVP 903
Db 421 HLVGTMTKIRVAARSNGGVSHGTSSEVINQDTLEKLAQGNESFLYGLINKSVP 480
Qy 904 VIVIVALLIIPVVIITACYWRNSRSGDKORSFKINDGSGVHMASNNLWDAQNPQN 963
Db 481 VIVIVALLIIPVVIITACYWRNSRSGDKORSFKINDGSGVHMASNNLWDAQNPQN 540
Qy 964 MYWTAGWMTNNRNGQALYSLTPNAQDFWNCDDYSGTHMRPGSEHHTYAQLTGQGN 1023
Db 541 MYWTAGWMTNNRNGQALYSLTPNAQDFWNCDDYSGTHMRPGSEHHTYAQLTGQGN 600
Qy 1024 MSTFYGNQYHDPSPYATTTLVLSNQPAWLNKLRAPAMPTNFPVPEPARYAHTAG 1083
Db 601 MSTFYGNQYHDPSPYATTTLVLSNQPAWLNKLRAPAMPTNFPVPEPARYAHTAG 660
Qy 1084 RRSRSRSDSGGTLNGGLHRTSGSQSDSPPHITVSVQLHSDGSGSKERTGERT 1143
Db 661 RRSRSRSDSGGTLNGGLHRTSGSQSDSPPHITVSVQLHSDGSGSKERTGERT 720
Qy 1144 PPKNTLMDFTIPPPSPPPGGHVITATRQLNGRSTPREDTIVSDGAFARVQVNR 1203
Db 721 PPKNTLMDFTIPPPSPPPGGHVITATRQLNGRSTPREDTIVSDGAFARVQVNR 780
1204 PTKRNNLGLRPLAGKXDDQSRSSLMDDGGSSADGENSEGVDPRGGVRAVPMGI 1263
Db 781 PTKRNNLGLRPLAGKXDDQSRSSLMDDGGSSADGENSEGVDPRGGVRAVPMGI 840
Qy 1264 SASTLASCTGNTGATFRSIPRNGIVTQBT 1297
Db 841 SASTLASCTGNTGATFRSIPRNGIVTQBT 874

RESULT 3
P91572
ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN ZK377.3.
OS Caenorhabditis elegans.
OC Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RW [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thiers-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstein L., Wilkinson-Sproat J., Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nhan M., Hawkins J.;
RL Submitted (FE8-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (FE8-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U08183; A052658.1; -.
DR HSP: P6276; 11LK.
DR IYTERPRO: IF003006; -.
DR PIR: P00047; Ig; 4.
SQ SEQUENCE 423 AA; 46544 BW; 45340DB68D575E5 CRC64;

Query Match 32.5%; Score 2232; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.6e-151;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYTLGPHYTHHTHTYINFDKIPNASHAPVIEHPIDVVVSRGSPATLNCAGKSTAKI 60
Db 1 MYTLGPHYTHHTHTYINFDKIPNASHAPVIEHPIDVVVSRGSPATLNCAGKSTAKI 60
Qy 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKVNKNGKSDAGATYCVASNEGEVKS 120
Db 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKVNKNGKSDAGATYCVASNEGEVKS 120
Qy 121 NBGLSKLALREDFVRPRPTVQALGEMVLECSPPRGPFPPVSWRDKDLRIQMPNR 180
Db 121 NBGLSKLALREDFVRPRPTVQALGEMVLECSPPRGPFPPVSWRDKDLRIQMPNR 180
Qy 181 YTLHSGDLIDPVDSDSGTQVCANVMGERSVNPALSVFEPKFEQPKMTVDWG 240
Db 181 YTLHSGDLIDPVDSDSGTQVCANVMGERSVNPALSVFEPKFEQPKMTVDWG 240
Qy 241 AAVLFDCRVPDQPOITWRKNEPMPTVRIAYIAKNRGLAIEHVQSDSGEVTYARNP 300
Db 241 AAVLFDCRVPDQPOITWRKNEPMPTVRIAYIAKNRGLAIEHVQSDSGEVTYARNP 300
Qy 301 AGTLASAHLVQAPPSPQTKPADQSPAGTATFCTLVGQSPATFWSKQDQLLF 360
Db 301 AGTLASAHLVQAPPSPQTKPADQSPAGTATFCTLVGQSPATFWSKQDQLLF 360
Qy 361 STYSDAGRTK/SPGTQTLTIEEVQVDBGAYCAGHNSAGSLSKAALKATFETKRVQK 420
Db 361 STYSDAGRTK/SPGTQTLTIEEVQVDBGAYCAGHNSAGSLSKAALKATFETKRVQK 420
Qy 421 KSK 423
Db 421 KSK 423

RESULT 4
O44924

ID 044924 PRELIMINARY; PRT; 1395 AA.
 AC 044924;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE ROUNDABOUT 1.
 GN ROBO1.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98117249; PubMed=9458045;
 RA Kidd T., Brose K., Mitchell K.J., Petter R.D., Tessier-Lavigne M.,
 RA Goodman C.S., Fear G.;
 RT "Roundabout controls axon crossing of the CNS midline and defines a
 RT novel subfamily of evolutionarily conserved guidance receptors."
 RL Cell 92:205-215(1998).
 DR EMBL: AF040989; AAC38849.1; .
 DR HSP: P56276; 1TLX.
 DR FLTBASE: Fbgo0005631; robo.
 DR INTERPRO: IP0001777; .
 DR INTERPRO: IP0003006; .
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; ig; 5.
 DR PRINTS: PR00014; FNTYPEII.
 SQ SEQUENCE 1395 AA; 151778 MW; B820E23A5218983 CRC64;

Query Match 23.14; Score 1588; DB 5; Length 1395;
 Best Local Similarity 31.04; Pred. No. 1.7e-104;
 Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

QY 29 APIVIEHPIDVWVSRGSPATLNC--GARPSTAKITWKDQGVITNKGDSHVRILVDTG 86
 DB 55 SPALIEHPDVLVAKNEPATLNCVKGKPEPT-TEWKGDEGVSTNEM--SRHWQKDG 111
 QY 87 SLFLVNVSGKMGKDSAGATTCVASNEHGEVKSNEGSLKMLRDEFRVPRPTVQALG 146
 DB 112 ALFTPTMGK--KEQDGGYWCVAKNRVGQVSRASLIQVLRDQFVPRKDTVAKG 169
 QY 147 EMALVCEPPFPFPPVSVWRKD-----KELRIQDMPTLTHSGNLIIDVDRSDG 200
 DB 170 ETALLECPCGPIPEPTLWIDGVLDDKMGFASGSRVIRVGGNLSISWEPIDEG 229
 QY 201 TIYCVANNMGVSRVSNPARLSVFEKPFQEPKDMVTVGAAVLFCRVGDPGQPIQWK 260
 DB 230 NYICIAQNVGSRSSYAKLVQVQKPFYKMEKQVLMVLGQATFHCVSQGVDPFVKLVK 289
 QY 261 RKNPEPMPTVY IAKDNRGLRIERVQDSBGYYTCARNPAGTLEASHLRVQAPSPQT 320
 DB 290 KEGNGIPVSRARILHDEKSLISNITPTDGTVCRAHNVNQI SARASLVHAPNFKT 349
 QY 321 KPADQSPVAGGTATFECTLVGQSPATVNSKEGQDLFPSTVSADGRTYSPGTLTIE 380
 DB 350 RPSNKGVLGVNLQVLPCHASGNPPSPSVFTKGVSTLMPN--SSHGQTVAAQDTLQIT 407
 QY 381 EVRQVDEGATVCAGNSAGSSLKAAKALTFTFKGVQRKQKSKMGKQKRVNSQIKYLI 440
 DB 408 DVHQDEGTYTCSAFSPVDSSTVRFVQVS----- 437
 QY 441 SAVTNGPAPKPPPTIEHQHQNQTLMVGSAILPQASQKPTPGISLKDGLPIDITDSRI 500
 DB 438 ----SVDERPPIIQIPANQTLKPGSAVLPCRAATGNPSPRIKVDEGHAQV--AGNRY 491
 QY 501 SQBSTGS LHIALDKPQGVYTICIAKNEDSGSTWSALIVEDHSTNAQFVMPDPNFPNS 560
 DB 492 SIQSSLSRLVDQLGSDSGTYCTASGERGTSHAATLIVKPGSTS--LHRAADSTTPA 550
 QY 561 SPQIPIVNVYDTEVELN--NAPSTSGAGPIPIYIQQYSPLDQVFNIPDYVASTEY 618
 DB 551 PPGTPKVLNVRTSISLWAKSQKRGAGPIGITYVEYFSPDLQGMVAHRVGTQV 610

QY 619 RIKLPGSHSITMVRINEKEGIGTSPVSSALVITTSKPAQVALSDKNKODMAIEKRL 678
 DB 611 TISGLTPGTSTVFLVRAENTQISVPSGLSHVILKSVIKTIEADFDAAAN----DLSAARTILT 666
 QY 679 SEQLLKLLEVKYITNSTAVRLVKKKLL--BELIDGYIKNGRPFRTNDQY--VNVISPS 734
 DB 667 GKS-VDELIDASINSAVLEKMLVLSADEKIVYEGRIYK--DASVPASQYHSITVMDAS 724
 QY 735 TENVTVNMPFTNTEVFFVPIPHSGVSHIGAPNSMDVLEAEPSSLPEDVRKRNMLN 794
 DB 725 AESFVGNLAKITKYEFLTPP---PETIEGQPSNKTALTYEDVPSPAPPQITGMNQ 781
 QY 795 TTLRLSKWAPKADGINGILAGFQI-VITVQAPNMRNNRNTINERASVTLFLHVTGMTYKI 853
 DB 782 TAGWVRWTPPPSQHNGNLYGTKEVSGAGTMRKVLANKMTTSTVLLANLTTGAVTSV 841
 QY 854 RVAASNGQVSGVSHGTVIMQDLEKHL-----AQQNESFLGLINKLSHVP-- 903
 DB 842 RLASPTKAGQSPKPSILMD-PTHVHPRAHPSGTGDEHGOQLTTH--NNGNIPP 898
 QY 904 -----VIVTALLIFVVI IACTVWNSRSDGKDSFIKIND 942
 DB 899 GDNINPTTHAKTIDYLSGFWLWVCLVLDLVLSAISITVFKRKHQKTELGHLSVSD 958
 QY 943 GSVHMSNM-----LMDVAFQNPQMNPTAGMTMNRNGALISLNPQAQDFNMCDY 998
 DB 959 NZTALNINSEKSL-----IDHGHGNTADTDKDSGLSGESLKHSHVNSQSNTNSD-- 1011
 QY 999 SGTHRGSGSEHHTAQLTGGPGNAMSTPYG-NQYHDDPSPTATTLV----- 1045
 DB 1012 GGT-----DIAEV--DTRMLTTPYCKSPDNPTPATMTIGTSSSECTKTT 1058
 QY 1046 -----LSNQQA--WLNKMLRAPAMPT-----PVPE--PPARTAD 1079
 DB 1059 SISADKSGTSHPSIDAFAGPAPVPAVPSKYLQITVEFINWSELPFPPEPSPSTTG 1118
 QY 1080 HTAG--RRSSSRASDGR-----TLNGLHRTTSGSQR-----DSPPTDWSY 1122
 DB 1119 YAGSGPSSKSSSSAGSGISTNQSLANHSSSSGGFSAGVSPQATVACPNPNTVNS 1178
 QY 1123 VQLSSDGTGSSKERTGERTPNKTLMDFPPTPPSPNPPPGGSHVDTATRRQ----- 1175
 DB 1179 PLSAVAGGTQRTQITPTNQHPQLRAY-PATTGCGAVVP-NHL-PFATQRAASETA 1235
 QY 1176 -LNRGSPREDTYS-----VSDGAFARVDVNA-----RPTSRNRLN----- 1211
 DB 1236 GLAARCAQSPACNSDALATPSMPQPPPPVPEGVGVPVPHSPHNPHTPSNQIYQC 1295
 QY 1212 GRLGPLKXKDDOSQRSLLMDGGSSADGSENGDVP 1250
 DB 1296 SSECDSHRSRSGSHKRLQLEHGSSAKRGGRHRRAP 1334

RESULT 5

Q9W213

ID Q9W213 PRELIMINARY; PRT; 1395 AA.

AC Q9W213;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ROBO PROTEIN.

GN ROBO.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

[illegible]

RESULT 6
089026
ID 089026 PRELIMINARY; PRT; 1612 AA.
AC 089026;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DUT11 PROTEIN.
GN ROBO1 OR DUT11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Wu M.C., Lowe N., Pordham R., Rabbitts P.;
RT "The mouse homologue of human DUT11/H-robo1 gene: protein sequence and
RT chromosomal location.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y17793; CAA76850.1; -
HSP: P56276; 17LK.
MD: MGI:1274781; Robo1.
DE INTERPRO: IPR001777; -
DE INTERPRO: IPR003006; -
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; ig; 5.
SQ SEQUENCE 1612 AA; 176406 MW; 5F2988C544796B4B CRC54;

Query Match 21.94; Score 1505.5; DB 11; Length 1612;
Best Local Similarity 32.14; Pred. No. 1.7e-98;
Matches 402; Conservative 176; Mismatches 498; Indels 175; Gaps 34;

Qy 30 PVIEHPIDVVSRSGSPATLNCAG-PSTAKITWKDQGVINKEQVNSHRVLDTGSL 88
Db 29 PRIVHPDGLVSKGSPATLNCAGRPPTIEWYKGGSERVDTKDPSSRHLGSGSL 88
Qy 89 PLLKNSGKNGKSDAGATVCVANSHEGVEKSGSLKLAHDFVRPRTVQALGGM 148
Db 89 PLLKNSGKNGKSDAGATVCVANSHEGVEKSGSLKLAHDFVRPRTVQALGGM 147
Qy 149 AVLECSPPGPPPEPVSWRDKELRIQDMPTYLHSDGNLLIDVDRSDSGTQVANN 208
Db 148 AVLECSPPGPPPEPVSWRDKELRIQDMPTYLHSDGNLLIDVDRSDSGTQVANN 205
Qy 209 MGVRSNPNARLSVFERKFPQKDMTVGVGAUFLDCRVTDGQPIWKKRKNMPPV 268
Db 206 MGVRSNPNARLSVFERKFPQKDMTVGVGAUFLDCRVTDGQPIWKKRKNMPPV 265
Qy 269 TRAYIAKDNHGLRIERVQSGDEGYCYARNPACTLEASHLRVQAPSPQTKPADQSP 328
Db 266 SR-TLEKDHDTLIRKVTAGDMGTYTCAENMYKAEASATLTVQHPHFVPRDQVA 324
Qy 329 AGGTATFECTLVQSPPAYFWSKQGLDLPST--VSADGRKTVSPGTGLTIEVRQVD 386
Db 325 LGRTVTPQCATGNQPAIFWRRGSQLP-STQPQSSSRFSVSGQGLDITNVQRSD 383
Qy 387 BGATVCAGNMSGAGSLKAAKATFETGRVQKRSKMGKQKQKQSIKILYSATGN 446
Db 384 VGYICQLVAGSIIITKYLE-----VTDV 409
Qy 447 TPAKPPPTIEHGHQNTLMVSSAILPCQAGKPTGISHLRGLPIDTDSRISQSTG 506
Db 410 TADRPPTVIRGPPVQTVAVDGTLLSCVATGSPAPTLWRKDGVLVSTQDSRIKQLSG 469
Qy 507 SLHIAKDKPTGVYTCIAKNEGBSTWASLVEDHTSNAQVVRNPPSNFSPSTQPI 566
Db 470 WLQIRYAKLDTGRITCASTPSGSEATWSAYIEVBFQGVQVPPPTDPLSAPSKE 529
Qy 567 IWNVDITEVLEHNPASTSGAGPTIGYIQYSPDLQTFWNPVQYVATSEIRIKLPS 626
Db 530 VTDVSKNTVLSGPNLNSGATP-TSYIEAFSHAGSGSQWAEANVKITFAIKGLPN 588

Qy 627 HSTMFVIRAREHEGIGTIPS-VSSALVTSKPAQVALSDKNMDMAEAKRLTSEQLIK 685
Db 589 AITFLFVRAANATGISDPSISDPVTKDVPPTSQVDHKKQ-----VQRELGNVLHL 641
Qy 686 EEVTKTSTAVRLFWKKRLEELIDGYIKWRGPPPTN-DNQYV--NVTSPTENTYVSN 742
Db 642 HNPITLSSSSVEVMTVDQSQSYIQYKILYRPSGASHGSEMLFVFNRTKNSVTPD 701
Qy 743 LMPFTKEFFVPIPHSGVSHIGAPSNMDVLEAPPSLPEDVFRMG--NLITLIS 800
Db 702 LRAGVWTEAKRPP--FNEFGQSEIKPAKTLLEAPSPAPRPSVSKMDGMDGATLMT 758
Qy 801 WKAPFADGICILAGQIVIVQAPNNNNITNERRASVTLFHLVTGHTYKIRVAARS 860
Db 759 WQPPEDTQGMQVETKWCNGNETYKINKTVDSGSTSVPSLVPGLRYSVEVWASTG 818
Qy 861 GGCV-----SHGTSEVIMMDLEKHLAQKESFLYGLKNSHVPVIVAI 910
Db 819 AGGVTSPEQPIQLDSRGNVSPEDVQSLAQISDVVRPFIAGIAGACWI-----I 871
Qy 911 LILFVILIAITCTWNSRNSD-----KDRSE-----IKNDGSMVMSN--WLDVAQN 958
Db 872 LWFSTLWLYNHRKKNGLSTYAGIRKVPSTFTPTVTVTQRGGAIVSGGRGOLLNISE 930
Qy 959 PNQNPNTAGMTMNNRNGALISLTNPAQFNNCDSTGTMHRPSEHHYHTAQTLG 1018
Db 931 PATPPLADTWPNTGNHNDGNCINCTAGNSDSNLTTTS---RPACDIANYQLDN 986
Qy 1019 GPNAM---STFYC-----NQTHD-----DPSFTATTLV---LSN 1048
Db 987 KQTNMLPESTVYGDVLSKNIMKMTFNSNLDKGRFVNSQSPPTATTLQIANLS 1046
Qy 1049 QQPAMADKDLRAPAMTNPVPEPPARYADHAGRRSSSRSDAGNLT-----NGL 1102
Db 1047 NNNAGGAGSSKHWKPGQPEVAPITNYIMENKLNLDYRANO--TTPPTPTMYST 1103
Qy 1103 HRTSGGSRSDSPHTDSTVYLHSDSGTSGSKERTGERPT--PNKTM---DFIPPPP 1157
Db 1104 DQMTGSGTNSD-----RGSTSGSGKHKG-ARTPKAKGGGNMADLPLP 1151
Qy 1158 SNPP-----PGGHV---DTATRQLNRGISTP 1182
Db 1152 AHFPNNSNSETYNMSDESTQDMCPVPPAMPYLLQDELQEEDERGGPT 1202

RESULT 7
Q9Y6N7
ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6N7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Petter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls exon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL: AF040990; AAC39575.1; -
DR HSP: P56276; 17LK.
DE INTERPRO: IPR001777; -
DE INTERPRO: IPR003006; -
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; ig; 5.

RESULT 9
Q9QZ13
ID Q9QZ13 PRELIMINARY: PRT: 1060 AA.

QY 16 YINFDKIPNA---SNLPAPVIIHPIDVVVSRGSPATLNCGA----KPSTAKITWYKDQ 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 24 YLELSSPGSRVPDAMPRIIVEOPPDIVSVRGEPATLECPARGPRPN---IDWKNGA 80

Q1 68 PVTNKQDLSHRVILDTGSLTFLVKNVSGDQSDAGAYCVASMEGEVKNMGLS 127
Db 81 RVATAREDDPRHRLLLPSGALFFPPIVGRKRSR-PDEGVTTCVARNVLGAASNASLVE 139
Q1 128 AMLREDFVFEVPTVQALGGEVAMGESSPPRGPPFPVSMKRODKMLRITQDMPTLHSD 187
Db 140 AVLKRDQFQSPGNWVAVGEPAWKEVPPKGHPEPLVTWAKGKILK-BEESGRIIRG-C 197
Q1 188 NLLIDPVDSDSGPTVCVANNMGEVSNPRLSVFERKPEFGEKMDTVDGVAAVLFC 247
Db 198 KLMSSHTFKSQMGMTVCVANNMAGSGSGAEVLVLERSPFLARPINQVLADAVPTLC 257
Q1 248 RVTGDQQQITLTKKRNKEMVPTRAYIAKDNRLGLIERVQPSDGEVTCVARNAGTLEAS 306
Db 258 EYQDQFVPLHKKRDKGELPAGR-TEIRSDHSLMDIQSDSEDTTTCVANSVGRAES 316
Q1 308 AHLRVAPPSPFOTKPAQDSVAGGATFECTVLQGPSAPVAFVSKQQQDLFFSY-VSAD 366
Db 317 GSLVSHVQVPTKQKQVTPAAGVNSFQCECTKGNPFPALTWKRQDSQVLLFVSQSLQM 376
Q1 367 GRKVKVPTGTLIEEVQVQDEGATVCAMNSAGSSLSKAAALATFETKVRQKKKSMKG 426
Db 377 GRLVSPDQQLNITEVTKDGGVTCVCAVSGASIGLARALL-----ETIG----- 421
Q1 427 QKQKVSQSIKYLISAVTGNTPAKPPIETIEHQNQITLMVGSSALPCQASGKPTGISV 486
Db 422 -----NSTDG-----LPIILQGPANTLVGSSWLPGRVWNPQIQM 462
Q1 487 LRQGLPIDITDSRISQHSQTSGLHIAIDKKPDTGVTCIAKNDGEGTSWASLTVDHSTN 546
Db 463 KDKERLQGDQSGDNLMDNGLTHLHISQEDMGFGLSVKASSIIGRATWNSHRLKQDM-C 521
Q1 547 AQVRFMPDPSNPSPPTQIIVNWDTIEVLEHNAAPSTSGAGITGIIQITVPSDLQTV 606
Db 522 ASPGAPQSPNPPGPPQSPVITVETANSITLTKWPNQGA-TATSVIETRAQSAQNTW 580
Q1 607 FNIPDPTVASTERTIKGLPSSHVMFVIAENEKIGTIVSVSSALVTTSKPAQVSLKSN 666
Db 581 RTVADQVQLETTTISGLQPTIILYPLVYVAGANGLSEPPSVSEPVQDQSSLRPADPW 640
Q1 667 KMDMAJERKLSLEQLKLEWKTINSTAVLFWKKRLEELIDGVIYKRWGPPPTKNND- 725
Db 641 KGGRCGLA-----EVAVRQCEPTVLGRTLQVSWVDGVPVLQVGRVRSWAGLDQGSW 694
Q1 726 QTVNWSTPSTKNTVSNLMPPTINTEFFV-IPYHSVHSHGAPSNMDVLIAEAPSPSL 784
Db 695 TMLDQSGFKZSVLRLGPPGAQIQIKVQVQDQGL-----GASFPVTISPEEAPSGP 750
Q1 785 EDVRIM--LNLTLAISWAKPAKDGINGILKQFIVGQAPNNRNITITERAAVSTL 842
Db 751 QGVAVALGDQNSVTSNPPFLPQKNGVITBGTWCLGNESRFLHNSAGMARSVTF 810
Q1 843 FHLVITMTIKRVAASNGVGVSHGSEYIMV-----QDTLEKHLAAQDEN 889
Db 811 SGLPLQYIALYVAATASAGVGA-SAPVVLQFPFPAEPGEVSEGLAERLAKRVL 868
Q1 890 BSFLYGLNLSHVEVPIVAILLIIFVVIIICVYWRNSNGDKRSPKINDGSEVMAS 949
Db 869 PAFLAGSSAAAG-----ALLGFCALYRQQRKQELSH-----ITAS 906
Q1 950 NNNDWAGNQPQNPMTIAGRWMTNNRNGALV-----SLTPNQAFFNNCDYS 999
Db 907 FATTPAVSPFSEDLGSSSSRPPMG-LGPAFVPLADSWHPPRSPSPAGRGSC----- 960
Q1 1000 GTMRHPSSEHYH-----AQLTG-GP-----GNAMSTTG--NQFH 1038
Db 961 -CPSNPPDDRTIYNAGISLYLAQTAGANASGSEPPVSTDVPGVELGTTHGFGQHS 1019
Q1 1034 DDPSPATITLVLVNSQPAWL-----NDKMLRAPA-MPT-----NPVPEPARTAD 1079
Db 1020 GDPSTWS-----QYAPPNSDSDSGARGGQKLLGQVQMSLWPEALPPPSCELS 1071

Qy 1080 HTAGRASRSRSGDRTGLANGLRHRTSGSQRSDSPPTDVSQVLHSDGSGKSKERTG 1139
 Db 1074 CPBGPEELKAGSSD-----LEEWCPVPPEKSH-LWGSSSGACWAPPA 1115

Qy 1140 ERTPT-----NKLMDFIPPPSPNPFP-----HYVDTATRLQNLNG-----STPREDT 1186
 Db 1116 PRDPTPTSTSTQGSATLTLSPDPQPPTDIPHLGMFRVPLGSPSPLSVSQALSS 1175

Qy 1187 YDS-VSDGAFARVDNARPT-----SRNNLGR-----PLGKRDGDSQSLDM 1231
 Db 1176 HDGSPVGLGAGPVLSYHASPSPVSTASSAGRRQVTCMTPLLGHGHRATIRKPKAL- 1234

Qy 1232 DDDGSGSEADGSESDWP 1250
 Db 1235 -----PYRRHSPGDL 1246

RESULT 11
 Q9VP26
 ID Q9VP26 PRELIMINARY; PRT; 859 AA.
 AC Q9VP26
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE CG5423 PROTEIN (FRAGMENT).
 GN CG5423
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BREKLEY;
 RX MEDLINE=20190606; PubMed=10731132;
 RA Adams M.C., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Annatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang C., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankooh C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Bensley E.M.,
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu M.E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-R., Iyegwan C.,
 RA Jallali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobery C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern R.P., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rehnert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).

DR EMBL; AE003586; AAF51388.1; .
 DR HSP: P56276; 1TLX.
 DR FLXBASE; FB0001328; CG5423.
 DR INTERPRO; IP0001777; .
 DR INTERPRO; IP0001006; .
 DR PFAM; PF00041; fn3; 3.
 DR PFAM; PF00047; lg; 5.
 FT NON_TER
 SQ SEQUENCE 859 AA; 93916 MW; 5CFD69D984101BF8 CRC54;

Query Match 18.24; Score 1249; DB 5; Length 859;
 Best Local Similarity 32.14; Pred. No. 1.3e-00;
 Matches 309; Conservative 151; Mismatches 351; Indels 152; Gaps 25;

Qy 30 PVILIEPDIWVVSRSPTLWAGAKPS-TAKITWYKQGVPIWKNRSHVLDGSL 88
 Db 1 PRIVEPDIWVPAHEPPTLWAGAKSPPTIYKQGVPL-----KILPGSHRLTLAGSL 57

Qy 89 FLKLVNSGKNGKSDAGAYCVASNEHGEVKSNGSLKAMLEDFRFPVHTQALGEM 148
 Db 58 FLKLVNSG-----RCANLDEFTLEPQWTRIAQST 89

Qy 149 AVLECSPPGPGPEVPSVNRKDKELRQIMPTLTSLGNDLIPDVSBSGGTQCVANN 208
 Db 90 ALLECAAPGPIPEPTVWKGKGLDLEGSVRVIVDGNLAIQDAQTQDGTQCTAKN 149

Qy 209 MUGVRSNSPALSVEYKPFQEPKQMDVDVAGVDFCRVGTQDQPGQITWR--KNEM 266
 Db 150 PVQVRESSTLALVWVKPTITRGHGTQVLGASVFPFCRVGQDPMVDLWLTASGNN 209

Qy 267 PVTRYATIAKNGRLRIKVPQSDBEYTYCIARNPACTLEASHLVQAPPSQTKPADQS 326
 Db 210 PLDRVSLVD-LSLRLERVTIADSGEYSCADNNGVATAMGLTIVTAPPKIQPQASK 268

Qy 327 VPAGTATFECTLVQGPSATFWSBGGQDQLFPSTVSADGRTKVSPT-----GLTIEBVR 383
 Db 269 VELGADTSECAIKNGKPTFTWIKNNSTLIFGAPGLDFRPHSLNTEBGSILTLTRQ 328

Qy 384 QVDEGAY-CAGMNSAGSLSKAAKLTFTKGVQKKSKMGKQKQNVSIIRYLISA 442
 Db 329 RTDKQVLICLHMAEVSITSRVQ-----SLDGE----- 360

Qy 443 VIGNTPAKPPTTIEHGNQNTLVGSSALLPQASGKPTGSLWRLDGLDITDTRISQ 502
 Db 361 -----RPPPIISGPNVTLPIKSLATQCKALGLSPSTISWTRGIPQV-PSSKNI 412

Qy 503 HSTGSLHIAL-KKPTQGVTTICARNDEGSTWSALTYEDHTS-NAQVPMOPSPNPS 560
 Db 413 TTSQDILSLDLRQDGLITCVASSRAGKSTQWGLRELTPITPNKIFTRAPDQTKPS 472

Qy 561 SPQPIVNVITDVELHWNAPSTSGAGPIGTIIQYSPDLQGTWPNFQVATSEYRI 620
 Db 473 APOQKILNLTASALITVWPTSDKAGSSGLFVSEVYCTNQSRTWPIASRLSEPIPT 532

Qy 621 KGLAPSHSTMVIRAEHNGKIGTSPVSSALVTYS-----PAQVALSDKNKM 669
 Db 533 ESLTQGLQVIRAEHNSLGSFSPSPISPTAGKLVGVDRGSESTGTQSLLD----- 587

Qy 670 MAIAEKRLISDLKLEEVKTINSTAVFLRWKMKLEELIDGIYIKWGRPPRTNDQYVN 729
 Db 588 -----VELLQANDVLELAEASDSTARLSDIDS-QYIEGFIYLA--ELHSEKMK 640

Qy 730 VT-----SPSTENVSNLMPFTNVEFFVPIYHSGVSHGAPSNMDLTAAPPSP 785
 Db 641 VILLNKGQGLSCTPVLAKASTYEFFVFPYK-----SIVKGSNRMRLDVPFAP 697

Qy 786 DVIRIMLNTLRLSNKADKADI-NGLKGQIVQVQAPNN-----RNITTNEAASV 840
 Db 698 GMEADQFRTSVLWKLPPQKPTNRGLISVNVVYKGLDWNHTFRKIMTIDAAAPLT 757

Qy 841 TLFLHVTGTMKIRVAERSNGVG-----VSGTSEVIMNQTLLEKHA 884
 Db 841 TLFLHVTGTMKIRVAERSNGVG-----VSGTSEVIMNQTLLEKHA 884

```

Db 758 LLANLTGVTYYTIIAAVAATRVGVPPSKPAVLRIDARTQSLDTGYTRYPISRDADFL- 816
Qy 885 AQQENESFLTGLINKSHVPIVIVAILLIFVYIIACCYWRNSRNSGDKDRSFIAKIND-G 943
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 817 ---TQTWTVLLGS---IIAIVFLLGALVLF-----KRYQPIKQTSLG 854
QY 944 SVH 946
      : :
Db 855 SLH 857

```

DR PFAM: PF00041; fn3; 1.
DR PFAM: PF00047; iq; 5.
DR PRINTS: PR00014; FNTYPEIII.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN1.
SQ SEQUENCE 823 AA: 89715 MW: 36FC0B91F36F2F19 CRC64:

Query Match 16.3%; Score 1120.5; DB 5; Length 823;
Best Local Similarity 30.3%; Pred. No. 1.9e-71;
Matches 263; Conservative 129; Mismatches 296; Indels 181; Gaps 18;

```

SSU17_12
Q9VQ10
ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
CS C5481.PROTEIN.
CS C5481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
NX NCBI_TaxID:7227;
OR
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE:20960006; PubMed:10731132;
RA ADAMS M.D., Celinkier S.E., Holt R.A., Evans C.A., Cocayne J.D.,
Ananatzides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Rutton C.G., Wortman J.R., Vandell M.D., Zhang C., Chen L.X.,
Braddon R.C., Rogers Y.-H.C., Blazej R.G., Champs K., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Abganyai A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Baillet M.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brookstein P., Brottier P.,
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra L.,
Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
de Pablos S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Duran-Rocha S., Dunkov B.C., Dunn P.,
Durbini K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glocke A., Gong F., Correll J.L.H., Gu Z., Guan P., Harris M.,
Harris W.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Houston K.D., Houston K.A., Howland T.J., Wei M.-H., Ithegumgar J.,
Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang T., Lin K.,
Liu X., Matteli B., McIntosh T.J., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshfiei A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.B., Nelson K.A., Nixon K., Nusskern D.R., Pacien J.J.,
Palazuelos N., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Sanders R.D.C., Scheeler P., Shen H.,
Shue B.C., Siden-Kiamis I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.B., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodgate T., Worley K.C., Wu D., Zhao X., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhang Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL AB003586; AAF51373.1; -.
DR HSP: P56276; ITLK.
DR FLYBASE: FBgn0031341; CG5481.
DR INTERPRO: IPR001412; -.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.

```

QY 37 IDVVVSGSPALMCGAKPS--TAKTIVKYGQPVITVQKNEQVSRILVDLQSGFLPKLYS 95
Db 1 MDTVFKPMDDPTTQCGAENPTPTIQWKGQDEL---KTQDGSRLMAGGLFLFKVH 57
QY 96 GKNKGSDAGAITCVASNEHGEVYKNSGSLKAMLEDREFFVRPTVQALGGEVAMLEWCSF 155
Db 58 SR--RESADGATYCAKNEFVARSNRATIQYAFIRDFEFLPANTVQCAVEALMECCA 115
QY 156 PRGSPFPPVYSNRKDDKRLIQKMPRTILHSDGNLIDPDSQSDTQCVANNWQGEVYS 215
Db 116 PRGSPFPPVYSNRKNGQTLWVSKRIRIVDGNLAIQEARQSDGRTQCVKNNVGTRES 235
QY 216 NPRLSVTEKPKFQEPKMTVDGVANFLDCVRGDPQIQWKKRKNEMPVTRIAYIK 275
Db 176 ATALFLKVFVFFPLIRGQMQITAVNGSVVQCGIGDGLDPLVLRKATSGSNMPPRHVL 235
QY 276 DNRGLAIEVRVQSDSEDEYVCYIARNPACTLEASAHLVQAPPSPQTKAQDSVPAGTATF 335
Db 236 EDRSLLDQVTELDMEGTCEADNANGVSTIGLTVAPFKPIVRKQLVEJGDEVL 295
QY 336 ECTLVQGPSPATFWSKGQDQLLFPSTVSADGRTKVST---GLTILIEVRVQDGAVF 391
Db 296 ECQANGPRPLTIVSWVSGNSLLDGT--RDSRGMVLTPEGRSVLSIAREADSGKV 353
QY 392 -CAGMNSAGSSLSK--AALKATPETKGRVKKKSKMGQKQKQVSIIRTLISAVTNP 448
Db 354 TCNALNAVGSVSSRTVSVSDQFEL--- 378
QY 449 AKPPETIEHGQNYTLMVGSNAILPQCAQSGKTPPGISWLADGLDITD--SRISQSTGS 507
Db 379 ---PPPILBQFVNGTLVPSILVTPCRLTGPVPVQSVWLDGPIQDQHEKRLSDAGA 436
QY 508 LHADLKL--POTGYTTCIAKKNDEGSTNSASLTVEHDS--NAQFVRMPDPNSPSPSTGP 565
Db 437 LTISSQREHEDTGLTVCASNRKGSNSGSLRLDPTNPNIWFFRAFLSLTGPFGPG 496
QY 566 IIVNVITDEVLNANPSTSGAGITGYIYISPDQQTWFNIPDVFASYTERIKLAG 625
Db 497 QMWKEGMSVLTSWTSRNSKGSSGLVGYEMFGNKNEIDGAWAGTVRQNTFTQTGLL 556
QY 626 SHSYMVLARENKGIQTSPVSSALVTSKPAQVALSDKNK--- 667
Db 557 GVNVTFLIARNSHGLSFLSPSEPIVNG---TVSSEHSTLMFSLIHYFDLSH 610
QY 668 ---MDVAIAEKRLLSEQLIKEVTINSTAVLEWK--- 702
Db 611 PQRYFNSGLDSEARSLISGQDVELSNASVDSSTMKLWQVCNRLTDSIAPIASHIA 670
QY 703 RKL---EELIDGYIKNR---GPRTNDNQYVWYSPS--- 734
Db 671 RHLIRASFLQMTINGKYVEGFYIARQLNFIVNPNAPVTSINPLLGISTSSASAS 730
QY 735 ---TENVSINLMPFTN 748
Db 731 ASALITKPNIAAGARDGETQSGGAPPLNTKYKMLTILANGGASGSCITLGLQVTL 790
QY 749 YEFFVIPHSYVHSIGAPNSMDVLTA 777
Db 791 YEFFIVPETY---SVGKSGNSRIARLE 816

RESULT 13

[illegible][illegible]

RESULT 14
Q9V4J9
ID Q9V4J9 PRELIMINARY; PRT; 2016 AA.
AC Q9V4J9;
DT 01-MAY-2000 (TremblRel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE CG17800 PROTEIN.
 GN CG17800.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter J.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes K., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan K.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina M.V., Mobarry C., Morris J., Mosherif A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith Z.,
 RA Spier E., Spradling A.C., Stapleton M., Strang R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 DR EMBL: AEO03841; AAF59271.1;
 DR HSSP: P40189; IBQJ.
 DR FLTBASE: Fbg0033159; CG17800.
 DR INTERPRO: IPR000267;
 DR INTERPRO: IPR001777;
 DR INTERPRO: IPR003006;
 DR PFM: PFM0041; fn3; 6.
 DR PFM: PFM0047; ig; 10.
 DR PROSITE: PS00144; ASN_GLN_ASE_1; UNKNOWN_1.
 SQ SEQUENCE 2015 AA; 222109 MW; 64ARDE3BB7BD0AB0 CRC64;

Db 488 VSYLNITS---VHANGDGLYKCIASKVGVA---EBSAKLVNGLPTIRQMEKAAIVAGE 541
 QY 148 MAVLECSPPFPPEPVSVWRKDKXELIQMDPTLHSDGNLIDPVR-SDSGTQCVCA 206
 Db 542 TLIVTC-PVAGTVPDSDIVNERDRALEN-RKQVFPNGLILLENVERNSDQATTCVA 598
 QY 207 NNMGERVSNPARLSVFEPK----FEQEPKIMTVGVAALFDCVR-TGDPQPTIKW 260
 Db 599 KNGQBSARGSLGVQVMVFPVLVPSFGS---AADVDIASANCYVFGDGLPEIRWS 654
 QY 261 KKEPM---PVTRIADNGLRIERVQSDSEGYTCYARNPAGTLEASARVQAP 315
 Db 655 LNSAPIVNGMGFTLVRLNKRITSLNIDNSAHFQVTKIATNPAGTSEVIAELQVWPV 714
 QY 316 PSFQTKPADQSPVAGTATFECTLVQSPATYFWSK---BSQ-COLLFPTSVSDAGRT 369
 Db 715 PWLILEPTIDAFQSGDAKVECKADGPKPVQVTKWAVGDTGSETKOLAKSONIREE--- 771
 QY 370 KVSPTGLTILEEVPQVDEGATYVAGNMAGSGSLKAALKATFETKGRVQKKSNGMKQK 429
 Db 772 ---ESTLIRVNQITQWTEYLTCEAINIGSGIS----- 801
 QY 430 KNVQSLIKYLSAVGTGNPKAPPTPIHNGQSLQVWSSALLPQASCPPTPGISWLRD 489
 Db 802 ---AVIMLSVQA-----PFEETKLNRQTARRGEPAVLQCEANGKXPICLWNN 848
 QY 490 GLPID-----ITDSRISQSTSGSLHADLAKPDTGVYCIANKDEGSEWTSASLIVE 541
 Db 849 NMLDLPKMDNNTTIREELSTGVMSLSIKRTSRDSALPTCVATNAGFSDASINQIV 908
 QY 542 DHTSNAQVPMPPDPSPSPQVLIVNTVTEVLEWHNASTSGAGCPTGVLIQYSPD 601
 Db 909 E---VPMPLATGLKSGRSVQLSWAQ-YDGNSPIDRTIEFKRS- 951
 QY 602 LGQWTFNTPDVA---STETRIKGLKSHSYVMVIRAEENGIGTSPVSSALVTISKPA 658
 Db 952 -RASGSEIDRIVPGHTEAQVQLSPATTYINIRVAEN-AIGTSSSEAVTIT---A 1005
 QY 659 QVALSKKMMMAIARLUETSLQLKLEVTINSTAVLFWKKAKLLEL---IDGYIK 715
 Db 1006 EBAKSP-----PQNVKPEV---NQTMVTVVPPPTNNGELIGITGVG 1048
 QY 716 WRQPPNTNDQVNTVTSPE---NYVSNLMPFNIEVPFVITSGVSHGAP-SN 769
 Db 1049 TK-LSNWSSIVFETINFTIEGKEKNEELNLRVITQTSVVI---QAFNKIGAGPLSE 1103
 QY 770 SMNVLTAAAPSLPEPQVIRMLNLTLRSLWRAPRAGDINGILKQVIVGAPWNN 829
 Db 1104 EEKQPTABGTSPQSPDTACTITLSQTRVNGVSPLEANGVITKTV---YAPSDN 1160
 QY 830 NITT---NERAASVTLFHLVTGMT-YKIRVAASNGGVG 865
 Db 1161 YDTRHRYKTKASSDPTVLGAKRYNTVMQVLAITAGGDV 1201

RESULT 15

Q3NBA1

Q1 Q3NBA1 PRELIMINARY; PRT; 2016 AA.

AC Q3NBA1;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE DSCAM PRECURSOR.

OS Drosophila melanogaster (Fruit fly).

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

CC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Schumcker D., Clemens J.C., Shu H., Worby C.A., Xiao J., Muda M.,

RA Dixon J.E., Zipursky S.L.;

RT "Drosophila Dscam is an Axon Guidance Receptor Exhibiting

Query Match 9.64; Score 658.5; DB 5; Length 2016;
 Best Local Similarity 26.04; Pred. No. 7.7e-38;
 Matches 229; Conservative 130; Mismatches 369; Indels 153; Gaps 33;

QY 30 PVIIERPIDVSVSNGSPATLNC-GAKPSTAKITVYKQDQPVITNKQVNSHRIVLDGTS 87
 Db 429 PVIIAQFQETMEFGPSVFLKCVAGGN-TPELSNEGLDKRIANDNRVQGVQTVGVNDV 487
 QY 88 LFLKLVKNGSKGSDAGATYCVASNEHGEVYSNEGSLKMLAMLEDPRVFRPTQVAGLE 147
 Db 147 LFLKLVKNGSKGSDAGATYCVASNEHGEVYSNEGSLKMLAMLEDPRVFRPTQVAGLE 147

RT Extraordinary Molecular Diversity.*;
RL Cell 101:671-684(2000).
DR EMEL: AF260530; AAF71926.1; -.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
SQ SEQUENCE 2016 AA; 222124 MW; 95CF95488F2AD36C CRC64;

Search completed: January 22, 2001, 12:53:03
Job time: 1984 sec

Query Match 9.6%; Score 658.5; DB 5; Length 2016;
Best Local Similarity 26.0%; Pred. No. 7.7e-38;
Matches 229; Conservative 130; Mismatches 369; Indels 153; Gaps 33;

Qy 30 PVIIIEHPIDVVVSGSPATLNC--GAKPSTAKITWYKDGQPVITNKQVNSHRIVLDGGS 87
Db ||| : : : | | | : | | : | | : | | : : :
Qy 429 PVIRQAQOETMEPPGSPVLKCVAGGNP-TPEISWELDGKKIANNDRYVQGVTVVNGDW 487
Db : : : : : | : | : | : | : | : : : : : : : :
Qy 88 LFLKLVNNGSKMGKSDAGAYTCVASNEHGEVKSNEGSLKAMLREDFRVPRPTVALGGE 147
Db : : : : : | : | : | : | : | : : : : : : : :
Qy 488 VSYLNITS---VHANDGGLYKCIASKVGVA---EHSAKNLVGLPTIRQMEKKAIVAGE 541
Db : : : : : | : | : | : | : | : : : : : : : :
Qy 148 MAVLECSPPPGFPPEPVSWRKDDKELRIQDMPTRIISDGNLIDPVD--SDSGTVQVCA 206
Db : : | : | : | : : : | : : : : : : : : | : | : | : |
Qy 542 TLIVTC-PVAGTPIDSIWNERNALPIN--RRQKVPFNGTLIENVERNSDQATTTCVA 598
Db : : : : : | : | : | : | : | : : : : : : : :
Qy 207 NNNVGERVSNPARLSVFEKPK----PQEPKDMTVDVGAALFDCVR-TGDPQPIITWK 260
Db | : | : : : | : : : : : | : | : : | : | : : |
Qy 599 KNQBGTSARGSLSEVQVMVPPQVLPSPSGES----AADVGDIASANCVVPGLPLEIRWS 654
Db : : : : : | : | : | : | : | : : : : : : : :
Qy 261 RKEEPM----PVTRAIANKNRLRIERVQPSDEGEYCYARNPAGTLESAHLRVQAP 315
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 655 LNSAPIVNGENGPTLVRLNKRISLNLNIDSLNAPRGVYCIATNPAGTSEYVRELQVNVV 714
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 316 PSFQIKPADQSPVAGGTATECTLVGQSPAPFWSK----EQ-QDLLPSTVSADGRT 369
Db : : | : | : | : | : | : | : | : | : | : | : | : : :
Qy 715 FPMVLEPTDKAFAGSSDAKVECKADGPPKPPQVTKKAVGDTPEYKDLKASDIRVE--- 771
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 370 NVSPGTGLTIEEVQVDEGATWCAGNAGSSLSKAALKATFETGRVQAKSKMGKQKQ 429
Db : : | : : : : | : | : | : | : | : : : : : : : :
Qy 772 ---BSTLVDNRKRTNKGTYLCEAINGISGLS----- 801
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 430 KNVQSIKYLISAVTGHTPAKPPPTIERGHQNTLMVSSSAILPCQASGKPTGISWLRD 489
Db : : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 802 ---AVIMISVQA-----PPEFTEKLNRARGEPATVCEAENGKPIGILNNMN 848
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 490 GLPID-----IDSRISQHSQSLHIALKPKPTGVYTCIAKNDEGSTWSASLTVE 541
Db : : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 849 NNKLDLPKNDNWTTHREILSTGVMSLSLKRTERSDALFTVATNAPGSDASINMIVQ 908
Db : : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 542 DHTSNAQFVMKPPQSNPPSSPTQPIIVNVTDIEVELHWNAPSGAGPTGTYIIQYISPD 601
Db : : | : : : : | : | : | : | : | : : : : : : : :
Qy 909 E-----VPEMTALKVLDKSGSVQLSWAQ-PYDGNPLORTIIEFKRS- 951
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 602 LGQVFNIPDQVFA---STERYIKLQPSHSTFVIRAEKNEGKIGTPSVSSALVTISKPA 658
Db : : | : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 952 -RASWSEIDRVIVPGHTTEAQWQLSPATTNIRIARN--AIGTSQSSAEVITIT---A 1005
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 659 QVALSDKNKMDMAIEAKRLTSEQLIKLEEVKTNINSTAVLFGKKRKLLEL---IDGTIK 715
Db : : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1006 EEAQSGK-----PONIKVEPV---NQITMRVTKRPPPTKNGKGLIGTVG 1048
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 716 WRGPPPTNDNQVNVNVTSPSTE-----NYVSNLMFPTNTEFFVITPISGVSHGAP-SN 769
Db : : | : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 1049 TK-LSTWSSVTFETINFITEGKEHNLQNLRYVTQSVI---QAFNKIGAGPLSE 1103
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 770 SMDVLTAEPASLPPEDEVIRIMNLNLTIRISWAKPADKINGILKGFQIVIVGAPNNNR 829
Db : : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1104 EEKQPTAEPTSQPPSDACTTLTSTQIRVGVSPPLESANGVITKTV---YAPSEW 1160
Db : : : : : | : | : | : | : | : : | : | : | : | : | : |
Qy 830 NITT-----NERAASVTLFHLVTGMT-YKIRVAARSNGGVG 865
Db : : | : | : | : | : | : | : | : | : : | : | : | : |
Qy 1161 YDETKRHYTKASSDQVHLGLKRTNYTMQVLATTAGDGV 1201

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:18:45 ; Search time 233.01 Seconds
(without alignments)
242.281 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKMKHVPFLWMISLLSLSPN.....VLGGYGERGDNNLEERTES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Aligned: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	8724	100.0	1651	20 Y13566	Human Robo 1 polyp
2	8704	99.8	1649	20 Y08404	Human ROBO1 protein
3	3736.5	42.8	753	20 W83927	Human 785 protein.
4	1592	18.2	1395	20 Y13563	Drosophila Robo 1
5	1592	18.2	1395	20 Y08401	Drosophila sp. ROB
6	1500.5	17.2	1297	20 Y13565	C. elegans Robo po
7	1500.5	17.2	1297	20 Y08403	C. elegans ROBO pr
8	1498.5	17.2	1380	20 Y08402	Drosophila sp. ROB
9	1498	17.2	1381	20 Y13564	Drosophila Robo 2
10	913	10.5	434	20 Y13567	Human Robo 2 polyp
11	913	10.5	434	20 Y08405	Human partial ROBO
12	856	9.8	985	20 F41716	Human PRO860 prote

13	761	8.7	1257	20 W74152	Human LI cell adhe
14	708.5	8.1	1571	19 W42087	Human Down syndrom
15	707.5	8.1	1910	19 W42086	Human Down syndrom
16	702.5	8.1	1192	19 W57900	Protein of clone
17	700.5	8.0	1299	21 Y40439	Human Nr-CAM prote
18	691	7.9	148	20 Y13568	Mouse Robo 1 polyp
19	691	7.9	148	20 Y08406	Mouse partial ROBO
20	682	7.8	1728	12 R13144	Deleted in Colorect
21	681	7.8	1304	19 W59994	Human neural cell
22	667.5	7.7	1447	16 W68553	Deleted in colorect
23	667.5	7.7	1447	20 Y33498	Human DCC protein..
24	664.5	7.5	1028	19 W29667	Homo sapiens DL185
25	644.5	7.4	3117	21 Y53667	Sequence gi/332818
26	596.5	6.8	1018	17 W87028	Human contactin..
27	595.5	6.8	1018	15 W63759	Human contactin (E
28	581.5	6.7	1911	16 W71726	Human PTP-OB. Hom
29	581.5	6.7	1911	18 W27225	Human protein tyro
30	581.5	6.7	1911	20 W94027	Human protein tyro
31	570	6.5	1018	18 W06485	Rat contactin liga
32	569.5	6.5	4412	21 Y53666	Sequence gi/101742
33	568	6.5	1242	19 W52287	Rattus norvegicus
34	556	6.4	1496	20 W81030	Melanoma associat
35	556	6.4	1496	21 Y70469	Human p53 target m
36	551.5	6.3	1225	19 W52289	Homo sapiens cdo t
37	549	6.3	1897	21 Y81785	Human protein tyro
38	549	6.3	1897	21 Y56100	LAR tyrosine phosph
39	521.5	6.0	1139	19 W37779	Rattus norvegicus
40	515.5	5.9	1125	19 W52288	Rattus norvegicus
41	507.5	5.8	1251	19 W37778	Rattus norvegicus
42	506	5.8	1501	16 W72858	Rat receptor type-
43	505	5.8	1070	18 W08747	Human colon carcin
44	478.5	5.5	848	21 W88565	Human NCAM 140KD
45	474	5.4	2387	21 Y53665	Mechanical stress

ALIGNMENTS

RESULT 1

Y13566
ID Y13566 standard; Protein; 1651 AA.
XX
AC Y13566;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
PN W09925833-A1.
XX
PO 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97OS-0065543.
XX
PA (RECC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
DR WPI: 1999-338008/28.
DR N-PSDB; X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisured) polypeptide in contact with a cell expressing active
 CC Robo (roundabout) on its surface. The method comprises modulating the
 CC effective amount of Comm polypeptide in contact with the cell, where the
 CC amount of expressed active Robo is specifically modulated inversely with
 CC the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 CC cell. The method can be used to screen for agents that modulate Robo:Comm
 CC interactions. This is particularly useful for modulating nerve cell
 CC function.

XX
 SQ Sequence 1651 AA;

Query Match 100.04; Score 8724; DB 20; Length 1651;
 Best Local Similarity 100.04; Pred. No. 0;
 Matches 1651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWKHWPFLLWISLLSLSPNHLFLAQLDPDPEVERGNHGTPTPTSDNDNSLGYTCSR 60
 |||||
 Db 1 MWKHWPFLLWISLLSLSPNHLFLAQLDPDPEVERGNHGTPTPTSDNDNSLGYTCSR 60
 |||||
 61 LQEDPPFPRVHSDLVSKGEPATLCKAEGRPTTIEWKYGSEVETKDDPSRSHM 120
 |||||
 Db 61 LQEDPPFPRVHSDLVSKGEPATLCKAEGRPTTIEWKYGSEVETKDDPSRSHM 120
 |||||
 121 LPLSGSLFFLAIVGGRKSRPDGVTVCVARNYIGAVSNASLEVALLDGFPNVDVM 180
 |||||
 Db 121 LPLSGSLFFLAIVGGRKSRPDGVTVCVARNYIGAVSNASLEVALLDGFPNVDVM 180
 |||||
 181 VAGEPAWMCQPPRGHPPTISWKKDGSPLDDKDERITIRGKLMITYRKSDAGKYC 240
 |||||
 Db 181 VAGEPAWMCQPPRGHPPTISWKKDGSPLDDKDERITIRGKLMITYRKSDAGKYC 240
 |||||
 241 VGTNNMGERESEVAELTVLERPSFKRPSNLAIVTDSAEFKCEARGDPVPTVWKDDG 300
 |||||
 Db 241 VGTNNMGERESEVAELTVLERPSFKRPSNLAIVTDSAEFKCEARGDPVPTVWKDDG 300
 |||||
 301 ELPSKRYEIRDHDLTKIRKVTAGDSYTCVAENMKGAEASALTVOEPHFVVPKQD 360
 |||||
 Db 301 ELPSKRYEIRDHDLTKIRKVTAGDSYTCVAENMKGAEASALTVOEPHFVVPKQD 360
 |||||
 361 VVALGRITVTCCEATGNQPAIFVRRREGSNLFTSPQSSPSSRSVQTGLTITNVQR 420
 |||||
 Db 361 VVALGRITVTCCEATGNQPAIFVRRREGSNLFTSPQSSPSSRSVQTGLTITNVQR 420
 |||||
 421 SDWGYTICQILNAGSIIITKALEYTDVIADRPVPIRGVQNVAVDGTFFVLSVCA 480
 |||||
 Db 421 SDWGYTICQILNAGSIIITKALEYTDVIADRPVPIRGVQNVAVDGTFFVLSVCA 480
 |||||
 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLQDGTGRTCIATSPGEATSWAYI 540
 |||||
 Db 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLQDGTGRTCIATSPGEATSWAYI 540
 |||||
 541 EVQEGFVGPVQPPRTDNLIPSAPSKPEVTDVSRNTVLSWQPNLNGATPTSYLIRAFS 600
 |||||
 Db 541 EVQEGFVGPVQPPRTDNLIPSAPSKPEVTDVSRNTVLSWQPNLNGATPTSYLIRAFS 600
 |||||
 601 HASGSSQVIAENVKVTETSAIKGLKPAIYFLVRAANAYGISDPSQISDPVKTQDVLPT 660
 |||||
 Db 601 HASGSSQVIAENVKVTETSAIKGLKPAIYFLVRAANAYGISDPSQISDPVKTQDVLPT 660
 |||||
 661 SQGVNDKQVQRELGNAVLHNPVTVLSSSIEVHWTVDDQSQYIGYKILYRPSGANHGE 720
 |||||
 Db 661 SQGVNDKQVQRELGNAVLHNPVTVLSSSIEVHWTVDDQSQYIGYKILYRPSGANHGE 720
 |||||
 721 SDNLVFEVTRPAKNSVVPDLAKGVNYEIKARPFNFEGADSEIKFATLEAPSAFPQ 780
 |||||
 Db 721 SDNLVFEVTRPAKNSVVPDLAKGVNYEIKARPFNFEGADSEIKFATLEAPSAFPQ 780
 |||||
 781 GVTYSKNDGNTALLWSNPQPEDTQNGMWQYKWCVLGNTRYHINKTVGDSFVSVIP 840
 |||||
 Db 781 GVTYSKNDGNTALLWSNPQPEDTQNGMWQYKWCVLGNTRYHINKTVGDSFVSVIP 840
 |||||

Qy 841 FLVPGIRISVEVAASTAGSGVYKSEPOFTQLDAGHNVSPVEDQSLAQISDVVQKPAFI 900
 |||||
 Db 841 FLVPGIRISVEVAASTAGSGVYKSEPOFTQLDAGHNVSPVEDQSLAQISDVVQKPAFI 900
 |||||
 901 AGIGAACHILIMVFSILWRHRRKNGKLTSTYAGIRKVPSTPTPTPTVYQGRGAVSSG 960
 |||||
 Db 901 AGIGAACHILIMVFSILWRHRRKNGKLTSTYAGIRKVPSTPTPTPTVYQGRGAVSSG 960
 |||||
 961 RPLGLNISSEPAQMWLADTWPVTGNHNDSCISCTAGNGSDSLNLTYSRPACIAENY 1020
 |||||
 Db 961 RPLGLNISSEPAQMWLADTWPVTGNHNDSCISCTAGNGSDSLNLTYSRPACIAENY 1020
 |||||
 1021 WQLNKGQTNMLPESTVYGDVLSNKNIEKMTFNSPNLKDGRFVNPSQPTATTQILQ 1080
 |||||
 Db 1021 WQLNKGQTNMLPESTVYGDVLSNKNIEKMTFNSPNLKDGRFVNPSQPTATTQILQ 1080
 |||||
 1081 SNLSNNMNGSDGSGEKHWKPLQQQKQVAVPVQYIVEDQKLNWIDYRNDVTPPTIPWQ 1140
 |||||
 Db 1081 SNLSNNMNGSDGSGEKHWKPLQQQKQVAVPVQYIVEDQKLNWIDYRNDVTPPTIPWQ 1140
 |||||
 1141 STQNTQSTGTHSDRGSSSTSGSGKKGARTPAVKQKGNWALLPPPAHPPHSSNE 1200
 |||||
 Db 1141 STQNTQSTGTHSDRGSSSTSGSGKKGARTPAVKQKGNWALLPPPAHPPHSSNE 1200
 |||||
 1201 EYNIISVDESTDQPCMPVPMYIQQDELEEDERGPTTPVGAASSPAVSYSHST 1260
 |||||
 Db 1201 EYNIISVDESTDQPCMPVPMYIQQDELEEDERGPTTPVGAASSPAVSYSHST 1260
 |||||
 1261 ATLTPSQEELQPMQLQDPEETGHQHQDPRRRQPVSPPPPRPISPHYTIGISPLVS 1320
 |||||
 Db 1261 ATLTPSQEELQPMQLQDPEETGHQHQDPRRRQPVSPPPPRPISPHYTIGISPLVS 1320
 |||||
 1321 DMDTAPDEEEDAEADNEVAKMTRLLRLGLETPASSVGDELSVTSGMWNGSASEE 1380
 |||||
 Db 1321 DMDTAPDEEEDAEADNEVAKMTRLLRLGLETPASSVGDELSVTSGMWNGSASEE 1380
 |||||
 1381 DNISSGRSVSSSSDGFDDFAQVAAAAAYAGLKVARRQMDAAGRHRFASQCPRP 1440
 |||||
 Db 1381 DNISSGRSVSSSSDGFDDFAQVAAAAAYAGLKVARRQMDAAGRHRFASQCPRP 1440
 |||||
 1441 TSPVSTDSNGNAVMQKTRPAKLLKHQPHLRRETYTDLPPVPPPAIKSPTAQSKTQ 1500
 |||||
 Db 1441 TSPVSTDSNGNAVMQKTRPAKLLKHQPHLRRETYTDLPPVPPPAIKSPTAQSKTQ 1500
 |||||
 1501 LEVRPVVVKPLPSMDARTDRSDRGSSYKREVLGRQVVMVMTNPGDREAGQDNDQ 1560
 |||||
 Db 1501 LEVRPVVVKPLPSMDARTDRSDRGSSYKREVLGRQVVMVMTNPGDREAGQDNDQ 1560
 |||||
 1561 KGRGNAAKRLDLPKATHLQIEDILYPCRPFTPSNPNRSDSSSSMSRSGSGSRQEA 1620
 |||||
 Db 1561 KGRGNAAKRLDLPKATHLQIEDILYPCRPFTPSNPNRSDSSSSMSRSGSGSRQEA 1620
 |||||
 1621 NVGRNIAEMVLGYGERGEDNNEELETES 1651
 |||||
 Db 1621 NVGRNIAEMVLGYGERGEDNNEELETES 1651
 |||||

RESULT 2
 Y08404
 ID Y08404 standard; Protein; 1649 AA.
 XX
 AC Y08404;
 XX
 DT 24-JUL-1999 (first entry)
 XX
 DE Human ROBO1 protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 KW cell morphology; screening assay.
 XX
 OS Homo sapiens.
 XX
 PN W09920754.AL

XX
PD 29-APR-1999.
XX
RF 00-OCT-1998; 98W0-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
(REBC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; N08404.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
XX
Claim 1; Page 65-71; 80pp; English.
XX
This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1649 AA;

Query Match 99.8%; Score 8704; DB 20; Length 1649;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1649; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 KKKHVPFVLMISLISLSPNHLAQQLIPDPEDVERGNDHGPTPTSDNDNSLGYTGR 60
Db 1 kkkhvpfvlmislislspnhlqlipdpedvergndhgptptsdndnslgytgr 60
Qy 61 LRQDPPPRVIEHPDLSIVSKGEPATLNCKAGRPPTPIIEMKGGGEVETDKDPSRHM 120
Db 61 lrqdpfprvievhpdsilvskgepatlnckagrptptpiiemkgggevetdkdpsrhm 120
Qy 121 LLPSGSLFIRIVBGRKSPDGEVYVARNYLGEAVSHNASLEVAILLRDFGPPSDVM 180
Db 121 llpsgsllfirivbgrksrpdgevyvvarnylgeavshnaslevalldrfgppsdvm 180
Qy 181 VANGEPVWMEQPPRGHPPTIISNKKDGLDDERITIRGKLMITTTKSDAGKATVC 240
Db 181 vangepvwmecpprgheptislwkkdgldderitirgkklmitytrksdagkayvc 240
Qy 241 VGTNMGVERSEVAELVIERPSFVKRPSNLAVTDDSAEFKCEARGDPVPIVWKKDDG 300
Db 241 vgtnmgeresevaelvierpsfvrkpsnlavtdvdsaeffkceargdpvpiwkkddg 300
Qy 301 ELPSKSYREIDRTHLWIKVYAGDMGSTYCAENMVKGABASATLVQEPHFVYVPRQD 360
Db 301 elpsksyreidrthlwikvayagdmgstycaenmvkgabasaatlveqphfvvprqd 360
Qy 361 VVALGTRVTFQCATGNQPAIFWRRGSGNLLFSTQPPQSSSRPSVSQDGLITINWQR 420
Db 361 vvalgtrvtfqcatgnqpaiifwrrsgnllfstqppqsssrpsvsqdglditinwqr 420
Qy 421 SDWGTYICQLTVAGSITIKAYLEVTDVIADRRPPVIRGPPVNYVAVDGTFWLSCVATG 480
Db 421 sdwgtyticqltvagsitikaylevtdviadrrppvirgppvnyvavdgtdfwlscvatg 480
Qy 481 SPVPTILWAKDGLVLTQDSRIKQLENGVLIQIRYAKLGDGTGTCTIASTPSGEATWSAYI 540
Db 481 spvptilwkdglvltqdsrikqlengvliqiryaklgdgtgtctiastpsgeatwsayi 540
Qy 541 EVQEGFVPPVPPRPDNLIPSAKSPKPVTVDSNRNVTLSWQPNLNSGATPISYIEAFS 600
Db 541 evqegfvppvpprpdnlipsakspkpvtdvsnrnvtlswnqpnlnsgatpisyieafs 600

Qy 601 BASGSSWQVAVNKVETTSAINGLAPNAILFLVRAANAYISDPSQISDPVKVQDVLP 660
Db 601 basgsswqvavnkvettsainglapnailflvraanayisdpsqisdvpkvqdvlp 660
Qy 661 SQGVDEKQVQREGLNVLHNPVTLSSSSIEVHWTDQSQSYIGTKILYKPSGANJGE 720
Db 661 sqgvdekqvqreqlnvlhnpvtlssssievhwtdqsqsyigtkilykpsganhge 720
Qy 721 SDMLVFEVETPAKNSVVPVILAKGVNVEIKARPFNEFGADSEIKFAKLEAPAPQP 780
Db 721 sdmlvfevetpaknsvvpilakgvnveikarpfnefgadseikfakleapapqp 780
Qy 781 GVTYSKNDGNGTALLVSWQPPPEDTQNGMVEYKWCWLGNETRTHINKTVDSGTFVVP 840
Db 781 gvtyskndgngtallvswqpppedtqngmveykwcwlgnetrthinktvdsgtfsvvp 840
Qy 841 FLVPGIRYSEVAASTGAGSGVSKSEPOFIQDAGNVPVSPEDVSLAQIISDVVKQAFI 900
Db 841 flvpgiryssevaastgagsgvsksepofigdaghvpspedvslaqiisdvkwqafi 900
Qy 901 AGIGACWNIIMVYSIKLIRHKKRNGLTSTYAGIKVKVPSFTPTTPTVYRGCEVSSGG 960
Db 901 agigacwilmvysilrhrkngltstyagikrvkpsftptptvyrgceevssgg 960
Qy 961 RPTLLNISPEAQPWLADTWPMTGNHNDSCISCTAGNNGSDSLNLTSPADCIANY 1020
Db 961 rptllnisepaqpladtwpmtgnhndscisctagnngsdslntyspadcianya 1020
Qy 1021 WOLDNKQTKMLPESTVYGVDVLSKINENKTFNSNLDKGFVWPGQSYPTATQLIQ 1080
Db 1021 woldnkqtkmlpestrygvdvlskinemktnfnsnldkgfvwpgqsyptattqliq 1080
Qy 1081 SNLSNMNMGSGSGDEKHWPLQGGQKQVAVQYVIVEQNKIANDTANOTVPTPIPWQ 1140
Db 1081 snlsnmnmgsgsgdekhwplgggkqvavqyviveqnkianotvptpiipwq 1140
Qy 1141 SDYQNTGGSTNSDRGSSSGSGKHKARKPKVYKQGGNMWALLPPPHAPPPHNSSE 1200
Db 1141 sydqntggstnsdrgsssgsgkhkarkpkvykqggnmwallppphappphnsse 1200
Qy 1201 EYISVDESDTQEMPCVPVPMRYLQDELESEDEGRPTPPVPGGAASSPAVSTSHST 1260
Db 1201 eyisvdesdtqempcvpvpmrylgdeleeseedgrptppvpggaasspavstshst 1260
Qy 1261 ATLTPSQEELQMLQDCEPTEGHMQVDRRRQVSPVPPPPRISPHPTTGTISGLVPS 1320
Db 1261 atltpsqeelqmlqdcepteghmqvdrrrqvspvpppprisphpttgtisglvps 1320
Qy 1321 DMDTAPEEEDAEADMEVAKMTRILLRLGLEQTASSVGDLESSVTGSMINGWASSEE 1380
Db 1321 dmdtapeeedaeadmevakmtrillrlgleqtassvgdlessvtgsmingwassee 1380
Qy 1381 DNLSGSRSSVSSSGSFPTDADQAVAAAEYAGLKVARRQMGAAGRHHFASQCPRP 1440
Db 1381 dnlsgrssvsssgsfptdadqavaaaeyaglkvarrqmgaagrrhfhfascprp 1440
Qy 1441 TSPVSTDSNMSAAVMQTRPAKLLKHQPHLRRETYTDDLPPPPVPAIKSPQAQSKIT 1500
Db 1441 tspvstdsnmsaavmqtrpakllkhqphlrretytdlppppvpaiksptaqsqtg 1500
Qy 1499 LEVRPVVVKPLPSMDARTDRSSDKGSYKREVLGRQVDMRTNFGDPREAGQEQNG 1560
Db 1499 levrpvvvpklpsmdartdrssdkgsykrevlgrqvdmrtnfgdpregaqqng 1560
Qy 1561 KGRGNIAKAKRQLPKAKTHLQEDILPCYRPTFTPSNNPDRSSSSMSRSGSGRQJEA 1620
Db 1561 kgrgniaakrqlpkakthlqedilpcyrptftpsnnpdrssssmsrsgsgrqjega 1620
Qy 1621 NWCRMIAEMQVLGGYERGEDNNEELETES 1651
Db 1621 nwcrmiaemqvlgygergednneeleetes 1651
Qy 1619 NVGRRLAEMQVLGGYERGEDNNEELETES 1649
Db 1619 nvgrrlaemqvlgygergednneeleetes 1649

(claimed). They can also be used in screening assays, detection assays (e.g. chromosomal mapping, tissue typing, forensic biology), CC predictive medicine (e.g. diagnostic assays, prognostic assays, CC monitoring clinical trials, and pharmacogenomics), and methods of CC treatment (e.g. therapeutic and prophylactic) e.g. for neurological disorders.

XX
SQ Sequence 753 AA;

Query Match 42.84; Score 3736.5; DB 20; Length 753;
Best Local Similarity 99.44; Pred. No. 1.6e-188;
Matches 716; Conservative 1; Mismatches 0; Indels 3; Gaps 1.

Qy 57 TGSLLRQDEPPPRVIEHPSDLIVSKGEPTALNCKAERPPPTIEWYKGERVETDKDDPR 116
Db 18 sgsrlrqedfpprivehpsdlivskgeptalnckaerppptiewykgervetdkddpr 77

Qy 117 SHRMLLPGSGSLFLRLIVHGKSPDQEGYVVCVARNYLGEAVSHNASLEVALRQDFQNP 176
Db 78 shrmllpgsflflrlivhgkspdqegyvvccvarnylgeavshnaslevalrddfrqnp 137

Qy 177 SDVMVAVGEPVAVMECPQPPGHEPTTISWKSGPLDOKDERITAGGLMITYTRKSDAG 236
Db 138 sdvmvavgepavmcepprghepttiskwsgplddkderitirggkimitytrksdag 197

Qy 237 KTYCVGTNMYGERSEVALLTLVRSPFKRPSNLAVTDDSAEFCXARGEPPVTVWR 296
Db 198 kytcvgtnmvgersevaltlvrspfkpsnlavtddseefcxargeppptvwr 257

Qy 297 KDDGELPKSRTEIRDHTLKRVTAGMSGYSVCAENMWKGAESAATLTQW---PPPHF 353
Db 258 kddegelpkrsyrlrdhtlkrvtagsmsytcvaenmwkgaesatltqvyspphf 317

Qy 354 VVKPQDVALGRVTFTVQCEATGNPPAIFWRRRSGQNLFSFTQPOSSSRFSVSQTGDL 413
Db 318 vvkpqdvalgrvtvtfqceatgnppaifwrrsgqnlffsfyqpgsssrfsvsgtdl 377

Qy 414 TITNVQRSDVGYIICQTLNVAAGSITKAYLEVDIADRPVPIRGQPNVQTVANDGT 473
Db 378 titnvqrsvgyiicqtlnvagsitkaylevtadiadrppvirgqpnvqtvandgtfv 437

Qy 474 LSCVATSGSPPTILWRKDGVLVSTQDSRKQLENVLQIRTAKLGDTRYTCLASTPSE 532
Db 438 lscvatsgspptilwrkdgvlvstqdsrkqlengvlqiraklqdtgtrytclastpse 497

Qy 534 ATWSAYIEVQEGFVQPPRPDTPNLPISSAPSKPEVTDSRNTVTLNQNLMSGATPTS 593
Db 498 atwsayievqegfvppprpdpdpnlpiissapkspevtdsrntvtlnqnlmsgatpts 557

Qy 594 YIIEAFSHASGSSQWVAENVKYTEISAIKGLKPNAILFLVRAANAYISDPQSIDPVK 653
Db 558 yieafshasgssqwvaenvkytesaikglkpnailflvraanaysdpqsidpdk 617

Qy 654 TQDVLPTSGVDHKQVRELGNVLEHNPTVLSSSSIEVHWTVQQSQYIQCTKILVRP 713
Db 618 tqdvlptsgvdhkvrelgnvlehnpvtlssssievhwtdqgsyiqgctkilyrp 677

Qy 714 SGANHGSEDWLVFEVTRPAKNSVPIDLRKYNVYIKARPFNEFGADSEIKFAKTLEE 773
Db 678 sganhgsewlvfevtrpaknsvipdlrkynvyekarpfnefgadseikfaktlee 737

RESULT 4
Y13563
ID Y13563 standard; Protein; 1395 AA.
XX
AC Y13563;
DT
DT 30-JUL-1999 (first entry)
DE
DE Drosophila Robo 1 polypeptide.


```

RESULT 5
Y08401
ID Y08401 standard; Protein; 1395 AA.
XX
AC Y08401;
XX
DT 24-JUL-1999 (first entry)

```

RESULT 6
Y13565
ID Y13565 standard; Protein; 1297 AA
XX
AC Y13565;
XX

us-09-540-245a-18.rag

	QY	449	IADREPFPPGQGVNVTVDVGTFLVSCVATGSVPPTLLMRKDGLWSTQDSRIKQLENG	508
			: : : : : : : : : : : : : : : : : :	
Dd		447	tpakrppctiehgqntlmvgssallpcqsagkptgjswldrglpiditdrisqstg	506
	QY	509	VLIQIRYAKLGGTGRITCIASTPSCGAETSWAYIEVFQEVGVQPVPPTDPLNISAPSKPE	568
			: : : : : : : : : : : : : : : : : :	
Dd		507	slihldlkpdtgyvtciakneedgetwsaalvethdstnaafvrmpdpnsfnpstgpi	566
	QY	569	VTDVSRTNLVKQWPLNLSGAPT--SVIIIFAFSHAGSSGWTHVAEMKWTEISAIGKLKN	627
			: : : : : : : : : : : : : : : : : :	
Dd		567	ivmrvtevelhwnapstsgagpltygilysygsdlgtwfnlpdyvasteyrkqkgks	626
	QY	628	AIZYLEVARAANYSIDSPSOISDPFKT-----QDVLPISQGVGHKKVQREL-GNAVHLR	681
			: : : : : : : : : : : : : : : : : :	
Dd		627	hsymfiraaenektgypsvsalvttskaagavalsdknkdmaiaekrtseglkie	686
	QY	682	NPTVLSSSIIEHVMTVDQSQSYIQGYKILLR-PSGANSGSDMLVEFVRTPAKNSVIFPD	740
			: : : : : : : : : : : : : : : : : :	
Dd		687	eaktlnstefirfwkkrlkeelldyyikvwgprprrdnq-----yvnnfstpsenyvsn	742
	QY	741	LKRGVNTIKARFP-----FNFGQADSEIKFARTLEEAPPGQVTSKNDGNGTALIVS	797
			: : : : : : : : : : : : : : : : : :	
Dd		743	lmpftneyffiyphsgvhislgapsasnvdltaseappipepdvirml-nltrlriss	800
	QY	798	WQPPFEDTGQNMQVKEKWCNLNETRINKTVDGSTFSVIFPLPWGIRISVESAAGST	857
			: : : : : : : : : : : : : : : : : :	
Dd		801	wkapkadgingllkgifrsivvgagnpnncritnaeasvfihlvgtmytkirvaarsn	860
	QY	858	AGSGVSEPQFIQLDAHGNPVSPEDVSAQAQISDVVWKPAFTIAGAACWI-----I	910
			: : : : : : : : : : : : : : : : : :	
Dd		861	ggvyg-----shgtsemvmcqltkehlaaqeneesflglylinksbyrvivaloi	910
	QY	911	LMVFSIWLTERRHKKNGLISTYAGIRKVPFFTPPTVTYORGEAVSSGPRGLNISE-	969
			: : : : : : : : : : : : : : : : : :	
Dd		911	lilfvilylviaycyvnrsrsd----gkdrsf-----lkindsgvshman---nlwdvga	958
	QY	970	PAAQFWLAIDTFMPTGNHNHDCISCTCAGNAGSDSLNITTS-----RPACIANTNNLOL	102
			: : : : : : : : : : : : : : : : : :	
Dd		959	pncpymyntargtmnmnrngalylslpnagdftnacdcydgshmrpgsehhyhaigtq	101
	QY	1026	RQTNLMLPESTYVGDLNLNKINEMKTIPSNPLMKDGRGVNSQSQTPYTATQLQSLNLS	108
			: : : : : : : : : : : : : : : : : :	
Dd		1019	ggpan--stfyg-----ngyh-----dpspyattvlvisn-----	104
	QY	1086	NNNGSGSDSGKHMKPLQOKQEVAVPQYINVBQMKLKNDRANDVIPPTPYNQSTDQN	114
			: : : : : : : : : : : : : : : : : :	
Dd		1049	-----gg-----pawlin-dxmlrapamptnvpvepp-arayah	108
	QY	1146	TGGSYNSSDRGSS-----TSGSG-----GHKK--G	116
			: : : : : : : : : : : : : : : : : :	
Dd		1081	tagrrsrsrsrdgrtloglhrtsgsrdsrpsdhtdvsyvgllhsdgtdgskertge	114
	QY	1159	ARTPKYKGVGMNNWALLPPPAPHPPPNSESSEYNISVDESVDQEMCPVPVPMAYMQD	122
			: : : : : : : : : : : : : : : : : :	
Dd		1141	rtrp-pnktilm--dfippensppp-----pgghvrb--d	116
	QY	1229	ELEEDEEERGPTT 1241	
			: : : :	
Dd		1170	tatrrolrestp 1182	

```

RESULT 7
Y08403
ID Y08403 standard; Protein; 1297 AA.
XX
AC Y08403;
XX
DT 24-JUL-1999 (first entry)
XX
DE C. elegans ROBO protein.
KW
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function

```

PD 29-APR-1999.

[illegible][illegible]

```

RESULT 9
YI3564
ID YI3564 standard; Protein; 1381 AA.
XX
AC YI3564;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo-2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W09925833-A1.
XX

```

Db	526	svtlswtrkvgsslgvlylmafignkdvgwagvtrvqntftqtqlpgwmyffil	585
QY	635	RAANAYGSDSPQISDPWTKQDLPQSPGVDRKQVRE-LGNAVLHLNPTVSSSSIEV	693
Db	586	raenshglvysmpsepl-tvgrtyfngsidsearasilsgdvelsmasvdsmskl	64
QY	694	HWTDQOSQSTQGTKILYR-----	712
Db	645	twgi-inkgykyevfyarqplpnpvnpapvtsnplglgstasasasalistk	703
QY	713	-----PSGANHGSDNLWTFEVKTP-----AKNSVIPOLRKGVNTEIKARPF	754
Db	704	pnaiaagkrdgetasggggaptlntkyrmltillingasscsctitlgvlyteiffv	763
QY	755	PNFGADASEIKFAKLEEAPASQPGVTSVKNDGQTAILVSWQPPEDTQNGWQETK	814
Db	764	kyvesqkpsnsrlartledvpeapyeamall--nsavflwkapekldrhgvlvlnh	821
QY	815	WVLGNLETRYII-----NKTVDGSTFSVPIPLVGIRISYEVAASTGAGSGVSEQPI	869
Db	822	virgldtahfsrlitvtdiaaspvlilantetgmytvgvaagnnagvygpcvatl	881
QY	870	QLDARGNVSPEDVQAQSIIDVVKPAPFINGAGACWILIMVFSINWLY-----RHRKKN	926
Db	882	rlidptkridp-fingchrhvditqgfillgallvalmsfgamvfrkhhmmqk	939
QY	927	GLSTIYAG-----IRKVPSTFTPTVYQRRGEA-----SSGG-----RPLGLNISPEAQP	974
Db	940	salotrnrhndsdvklmels-----arqncwldstscmwmrpsoudlelmsk	992

QY 975 WLADWPVYGNHNDNSISCTGATGNGSDNSLNTYSRPDCACTANYNQLNDLQMLPMLP 1034
Db 993 hiaayvpcgpcgspaggtsgsgsgagsg-----asgddghghgsernqrqv----- 1044

QY 1035 STVYGV-DLSNKINEMKTFNSPKLKGDFVPSGQPTPYATTLQISMLSNMNM----- 1089
Db 1045 -----gvsnlptdeysfkgapeyryhgas--papaytsilspghggggggpryq 1098

QY 1090 -----GSDSDSGHXHWLQQQKEVAPVQVINBQNLKNDTRVNDTPPTIPNQSG----- 1141
Db 1099 grqvysyigqpmh-phyqqqgh-----gggaagtlqhqghaigqhqglpsnlygmstt 1153

QY 1142 -----YDQNTGGS-----YNSSD-----RGSSTSGSGQ- 1164
Db 1154 seipytytgprvsysegypqqdkgrhihtenksnchtysaapgakgsspsisqfas 1213

QY 1165 -----HKGARTPVPKQGMNMWADLL-----PPPPAR 1192
Db 1214 vrrqlppncsigresarfkvltadgknqplldtggssmcyogldsgcgsspspmn 1273

QY 1193 PPFHNSSEYNTISVDESVDQEMCPVPFARMYLQDELEEDERGPCTVPRGAASPA 1252
Db 1274 lmsdehahlytadgldld-----merlykvdeqpgggqqqlipvpqhpaeghl 1326

QY 1253 VYSYSHQSTATLTPSQEELQPLQDCFEETHGMHQRP 1289
Db 1327 qswrastrsrkqoe-----cikenseliyap 1355

```

RESULT 10
Y13567
ID Y13567 standard; Protein; 434 AA.
XX
AC Y13567;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 2 polypeptide.
XX
KW Comn polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.

```

Qy 780 QGVTSKMDGSIHALVSWQPPEEDQTQGMVQYKWCGLNETRYHINRTVDGSFSWVI 839
|||
Db 280 vsrtgkx----- 286

Qy 840 PFLVPGIRYSVEAASTGAGSGVKSEPOFIQLAHGNPSPEDQVSQAQQISDVQKPAF 899
|||
Db 287 -----pq----- 288

Qy 900 IAGIGAACWILMVFSIWLYHRKKNCNLSTYAGIKRVFSFTPTTIVTYQRGGEAVSSG 959
|||
Db 289 -----kng----- 292

Qy 960 GRPLGLNISEPAAQWLADWTGNWNHDCSICCTAGNCSNDLSNTIYSRADCANI 1019
|||
Db 293 -----stvan----- 297

Qy 1020 NNQLDNKTQILNMELPSTVTGDVDSKNINKEMTKTNSPLMKDGRFVNPSQPTPIATOLI 1079
|||
Db 298 ----- 297

Qy 1080 QSULSNMNNMSGDSGEKHKLQGCGQEVPAPQYINIVBQNKLWKDYFRVTPPTIPTYN 1139
|||
Db 298 ----- 297

Qy 1140 QSTDQNTQSTSNDRSGTSGSGHKRGARTPKVPQGGMNWADLPPPPPAPHPSNS 1199
||| ||| |||
Db 298 -----vp-----lppppqplgtel 313

Qy 1200 EETYNISVDES-YDEQPCPVPPARMYLQ--DELEEDEDERGPTPVVGGAASSPAANVS 1255
|| | : : : : : ||| ||| ||| : : : : : ||| ||| : : : : : ||| : : : : :
Db 314 ehvayeqgenysdsdcwcpplqvqlylhqledeled-eeddrrvtptprvgassp-aisf 371

Qy 1256 SHQSATLITPSQOEELPMQLQCDEETHGMQHQPDRRRQPVSPPPPRIPSPHTYGIS 1315
||||| ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 372 gggstalltspreemgmigasp----- 395

Qy 1316 GPLVSDMTDAPEEDEDAMEVAKQTRRLLLGLEOTPASVGDLESSVTGSMNGWG 1375
|||
Db 396 ----- 395

Qy 1376 SASEENISSGRSVSSDGDFDTAFDAQAAAAAYAGLVARRMQDAQAGRHHFAS 1435
|||
Db 396 -----xtss 400

Qy 1436 QCPRPTSPVSTDSNMSAAVMQKTRPAKLHWGP 1469
| | | | | | | | | | | | | | | | | | | | | |
Db 401 qrrprtspfstdntsaaLsgscrprrkhhqq 434

```

RESULT 11
Y08405
ID Y08405 standard; Protein; 434 AA.
XX
AC Y08405;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human partial ROBO2 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
XX
PN W09920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062821.

```

XX
PA (RBGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI: 1999: 312615/26.
DR N-PSDB: X57254.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 72-73; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 434 AA;
Query Match 10.5%; Score 913; DB 20; Length 434;
Best Local Similarity 23.2%; Fred. No. 1.2e-40;
Matches 258; Conservative 54; Mismatches 118; Indels 684; Gaps 12;
QY 360 QVVALGRITVFQCEATGNPQPAIIFWRBBSQNLFFSYQPSRSSFSVSGTDLITITWQ 419
Db 1 qivaggrtvtfpcekgpqpavfwqegsqnllfpnqgqgpcscsvsptgdlititniq 60
QY 420 RSDVGYYICQLLWAGSIIITKATLEVTDIADRPVPIRQGPVMTVANDGTFLVCVAT 479
Db 61 rSdggyllqaltvagsilakaglevtdltdpppiliqgpanqlavdytallkckat 120
QY 480 GSPVPTILMRADGVLSVQDSRIKQLENGVLQIRIARLGDGTGRTTCTASTPSGEATWSAY 539
Db 121 gdpLpIswlkegftfprdrpraticqeggtlqknrlsdttytcratssgeasvsw 180
QY 540 IEVQEGFVPGVPPRPTDFNLIPSAPKPEVTDVSRNVTLSWQPNLANSAGPTTSTIIEAF 599
Db 181 ldvtesgatis-knydladlpggspkqvtdvtnsvlswqgtpgtlpassyileaf 238
QY 600 SHASGSQWTVARNKVTETSAIKGLKNATYILFVRANAYGISDPQSDPVTIKVDLPL 659
Db 239 sqvsnswgtvanhvkttlytrvgirptiylmvrain----- 277
QY 660 TSQGVQDHQVQRELGNVYLHLNPTVLSSSIEBVHWYVDQSQVITQCTIKLYRPSGANHG 719
Db 278 ----- 277
QY 720 ESDLVFVFRTPAKNSWVPLDLRKGVNYEIKARPFNFQGADSEIKFAKTLLEASAPP 779
Db 278 -----pk 279
QY 780 QGVTVSKNDNGTALYLSKQPPEDTGNQVQBYKVKWLGNETRTHINKVDGSTFSVI 839
Db 280 vstqtq----- 286
QY 840 PFLVPGIRYSVEVAAGTAGSGVKSEPFQFLDLAHNGVSPEDYQLAQQISDVYQKAF 899
Db 287 -----pg----- 288
QY 900 IAGIGACWILIMVFSIMLYRKHKKRKLGLSTYAGIRKVPSPPTPTVTYRGGEASVSG 959
Db 289 -----kng----- 292
QY 960 GRGPLNLISEPAQWPLADTWPNVTGNNGNDCSCTAGNCGNSDLTTSRPADCIANY 1019
Db 293 -----stwen----- 297
QY 1020 NNQLDQKNTNMLPESTVYGDVLSNKNEMKTFNPNLKDGRFVNPSPQPTFYATQLI 1079

Db 298 ----- 297
QY 1080 QSNLSNNKNGSGDSGEKHWKPLGQKQEQVAPVQVIVBQNKLNKDYRANDTVPTPTFN 1139
Db 298 ----- 297
QY 1140 QSTDQNTGSGNSDRGSSSTSGSQGHKAGRTKVKPGQGMNADLLPPPPHPPPHSNS 1199
Db 298 -----vp-----lpppvqplpgtel 313
QY 1200 REYNISVDES-YDEMPCPVPVPMYLOQ---DELEEEDERGPTTPPVGAASSPAVST 1255
Db 314 ehyaqegqenydyswcpplpqtylhgledel-eedddrvtpgvrgvass-alsf 371
QY 1256 SHQSTATLTPSQEELQMLQDCPEETGHQHQDPRRRQPVSPPPPPRTPSPHTYGIS 1315
Db 372 gqgstlrltspreemqplgasp----- 395
QY 1316 GPLVSDMDTDPEEEDDEADMEVAKMQLTRALLRGLBQTASSVGDLESSVTGSMINGW 1375
Db 396 ----- 395
QY 1376 SASEEDNLISSGRSSVSSSDSGFTDADFACAVAAAEYAGLKVARBQMDAAGRHFHAS 1435
Db 396 -----xtfss 400
QY 1436 QCPPTPTSPVSDSNWGAVMQMTAPRAKLKHPG 1469
Db 401 qrrptstfstdntsaalsagqrprtrkhhgq 434

RESULT 12

Y41716

ID Y41716 standard; Protein; 985 AA.

XX

AC Y41716;

XX

DT 07-DEC-1999 (first entry)

XX

DE Human PRO860 protein sequence.

XX

KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.

XX

OS Homo sapiens.

XX

PN W09946281-A2.

XX

PD 16-SEP-1999.

XX

PF 08-MAR-1999; 99WO-US05028.

XX

PR 10-MAR-1998; 980S-0077450.

PR 11-MAR-1998; 980S-0077632.

PR 11-MAR-1998; 980S-0077641.

PR 11-MAR-1998; 980S-0077649.

PR 12-MAR-1998; 980S-0077791.

PR 13-MAR-1998; 980S-0078004.

PR 17-MAR-1998; 980S-0040220.

PR 20-MAR-1998; 980S-0078886.

PR 20-MAR-1998; 980S-0078910.

PR 20-MAR-1998; 980S-0078936.

PR 20-MAR-1998; 980S-0078939.

PR 25-MAR-1998; 980S-0079294.

PR 26-MAR-1998; 980S-0079566.

PR 27-MAR-1998; 980S-0079663.

PR 27-MAR-1998; 980S-0079664.

PR 27-MAR-1998; 980S-0079689.

PR 27-MAR-1998; 980S-0079728.

PR 27-MAR-1998; 980S-0079786.

PR 30-MAR-1998; 980S-0079920.

PR 30-MAR-1998; 980S-0079923.

PR 31-MAR-1998; 9805-080105.
PR 31-MAR-1998; 9805-080107.
PR 31-MAR-1998; 9805-080165.
PR 31-MAR-1998; 9805-080194.
PR 01-APR-1998; 9805-080327.
PR 01-APR-1998; 9805-080328.
PR 01-APR-1998; 9805-080333.
PR 01-APR-1998; 9805-080334.
PR 08-APR-1998; 9805-0801049.
PR 08-APR-1998; 9805-0801070.
PR 08-APR-1998; 9805-0801071.
PR 09-APR-1998; 9805-0801195.
PR 09-APR-1998; 9805-0801203.
PR 09-APR-1998; 9805-0801229.
PR 15-APR-1998; 9805-0801817.
PR 15-APR-1998; 9805-0801838.
PR 15-APR-1998; 9805-0801952.
PR 15-APR-1998; 9805-0801955.
PR 21-APR-1998; 9805-0802568.
PR 21-APR-1998; 9805-0802569.
PR 22-APR-1998; 9805-0802700.
PR 22-APR-1998; 9805-0802704.
PR 22-APR-1998; 9805-0802804.
PR 23-APR-1998; 9805-0802767.
PR 23-APR-1998; 9805-0802796.
PR 27-APR-1998; 9805-0803336.
PR 28-APR-1998; 9805-0803322.
PR 29-APR-1998; 9805-0803392.
PR 29-APR-1998; 9805-0803495.
PR 29-APR-1998; 9805-0803496.
PR 29-APR-1998; 9805-0803499.
PR 29-APR-1998; 9805-0803500.
PR 29-APR-1998; 9805-0803545.
PR 29-APR-1998; 9805-0803554.
PR 29-APR-1998; 9805-0803558.
PR 29-APR-1998; 9805-0803559.
PR 30-APR-1998; 9805-0803742.
PR 05-MAY-1998; 9805-0804366.
PR 06-MAY-1998; 9805-0804414.
PR 06-MAY-1998; 9805-0804441.
PR 07-MAY-1998; 9805-0804598.
PR 07-MAY-1998; 9805-0804600.
PR 07-MAY-1998; 9805-0804627.
PR 07-MAY-1998; 9805-0804637.
PR 07-MAY-1998; 9805-0804639.
PR 07-MAY-1998; 9805-0804640.
PR 07-MAY-1998; 9805-0804643.
PR 13-MAY-1998; 9805-0805323.
PR 13-MAY-1998; 9805-0805338.
PR 13-MAY-1998; 9805-0805339.
PR 15-MAY-1998; 9805-0805573.
PR 15-MAY-1998; 9805-0805579.
PR 15-MAY-1998; 9805-0805580.
PR 15-MAY-1998; 9805-0805582.
PR 15-MAY-1998; 9805-0805689.
PR 15-MAY-1998; 9805-0805697.
PR 15-MAY-1998; 9805-0805700.
PR 15-MAY-1998; 9805-0805704.
PR 18-MAY-1998; 9805-0806023.
PR 22-MAY-1998; 9805-0806392.
PR 22-MAY-1998; 9805-0806414.
PR 22-MAY-1998; 9805-0806430.
PR 22-MAY-1998; 9805-0806486.
PR 28-MAY-1998; 9805-0807098.
PR 28-MAY-1998; 9805-0807106.
PR 28-MAY-1998; 9805-0807208.
PR 30-JUL-1998; 9805-0904651.
PR 11-SEP-1998; 9805-0100038.

XX (GETH) GENENTECH INC.

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX

DR WPI; 1999-551358/46.

DR N-PSDB; 234069.

XX

PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -

XX

PS Claim 12; Fig 77; 530pp; English.

XX

CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. 133891 to
CC 234338, and Y41695 to Y41774 represent polynucleotide and polypeptide
CC sequence given in the exemplification of the present invention.

XX

SQ Sequence 985 AA;

Query Match 9.84; Score 856; DB 20; Length 985;
Best Local Similarity 23.34; Pred. No. 3.3e-37;
Matches 350; Conservative 144; Mismatches 420; Indels 590; Gaps 55;

Qy 64 EDPPPRIVPEPSDGLVSKGEPAITLNCKARBPPTPIEWYKCGERVTOKDQPSHMLLP 123

Db 6 qdsppgilhwpgdglfqqgpparmcqsagppptirvlwngplsmvpgdph-----llp 62

Qy 124 SGLSPFLA-----IVHGKRSPODGVYCVARNYLGAHSVHNSLEVALADDFRQWPS 177

Db 63 detllilqpparghhdgqalstdlgyvtceasnrlgtavrgarlsavlredfgiqpr 122

Qy 178 DVMWANGSPVMEQPPFRKHEPTISWKKDGLDOKDERITIRGKMLITITRKSDAKR 237

Db 123 dmnavvqgftlccgppwghpctvswkdqplalqgrhtvsgsllmaraksdegt 186

Qy 238 YVGVTCVWNGRGESEVLAELVLERPSFKRPSHLAVYDSDAEKCEARGDPVPTVWNRK 297

Db 183 ymcvatsnaghsraarvsigepqdytepvellavrlg----- 221

Qy 298 DQGLPKRTEIRDOHTLKIRKVIADGMSYTCVAENMKWGAESATLVYQPPHFVVKP 357

Db 222 -----lenv-----tll----- 228

Qy 358 RDQVVALGTFTVPCBATGNQPAIFVWRRGSSQLLEPSYQSSSSRFVSQGLDITIN 417

Db 229 -----npdpa-----eg----- 235

Qy 418 VQRSDVGTITQTLNVAWSIITKAYLEVTDVIADPPPIVQGRPVNQTAVDQFTVSLCV 477

Db 236 -----pkp----- 238

Qy 478 ATGSPVTLIWRKDGVLVSTQDRIKQLENGVLQIRYAKLDTGRYTCIAST-----PSGEA 534

Db 239 -----rpavtlsvkvsqpaagcs-----ytalfrgtatpgqg- 271

Qy 535 TWSAIEVQSGGVFPVPPRPTDNLPIAFSKPEVTVSRNTVLSWQPLNMGATPTSY 594

Db 272 -----gap-----weellag----- 282

Qy 595 IIEAFSHAGSSGQTVAEWVKTTETSAIKGLKNAYILFLVRAANAGISDPQISDPVKT 654

Db 283 -----wgsa----- 286

Qy 655 QDVLPTSGVGHKQVORELGNVLHLNPTVLSSSIEVHWTVDDQSQVIGYQIKTYRPS 714

Db 287 -----elgg-----lhw----- 293

Qy 715 GANHGSQDLWFEVTRTPAKNSVIPDLRKGVNYEIKARPPFNEFGDASIKPAKILEEA 774

||| ||| :|| : :|

[illegible]

RESULT 13
W74152

ID W74152 standard; Protein; 1257 AA.

XX

AC W74152;

XX

DT 05-MAY-1999 (first entry)

XX

DE Human L

XX

KW Human L

KW nervous system development; nerve regeneration;

KW neuronal cell cohesive interaction.

XX

1

Homo sapiens.
 PN US5872225-A.
 XI
 PD 16-FEB-1999.
 XX
 PF 18-NOV-1994; 9405-0341843.
 XX
 PR 26-JUN-1992; 9205-0904991.
 PR 18-NOV-1994; 9405-0341843.
 XX
 PA (UTCA) UNIV CASE WESTERN RESERVE.
 XX
 PI Lemmon V;
 XI
 DR WPI; 1999-166719/14.
 DR N-PSDS; X01598.
 XX
 PT Human L1 cell adhesion molecule - supports neurite outgrowth and is
 PI involved in nervous system development and repair
 XX
 PS Claim 1; Fig 3; 45pp; English.
 XX
 CC This sequence is the human L1 cell adhesion molecule (L1CAM) of the
 CC invention. L1CAM supports growth of neurites in vitro and is involved in
 CC development of the human nervous system and in nerve regeneration. It is
 CC useful in vivo and in vitro experiments on nerve growth and
 CC regeneration. L1CAM mediates cohesive interactions of neuronal cells to
 CC each other and to extracellular matrix.
 XX
 SO Sequence 1257 AA.

Query Match 8.7%; Score 761; DB 20; Length 1257;
Best Local Similarity 24.9%; Pred. No. 4.3e-32;
Matches 283; Conservative 159; Mismatches 422; Indels 274; Gaps 45;

Qy	14	LLSLSPHLFAGLIPDDPEVRGNDGQPIPTSDNDNSLGTGSRGLQDFPRVHEH	73
Db	11	lllcep-----clliqipeeyegshvmeppvt-----eqsprllrvf	46
Qy	74	PSDLIVSKGEWATLNCAGAPPTPTISWKGKGEVETDKD-----DPRSHRMLPSSG	126
Db	9	psdlivskgewatlncagapptptiswkgkgevetsk-----dprshrmllpssg	101
Qy	127	LFFLAIVBGRSKSPDGGVTVCARNLGEAVSHNASLVALRDDPQNPSD-----VWA	182
Db	102	snf-----agrgfivqfasmkigtamsh-----eirimeagapkwetkpvpe	149
Qy	183	VGEPAWBCQPPHGHPEPTISWKGKSLDDKDERITI-AGGLMIIITTRAKD-AGKVT	240
Db	150	egesvlpncppsaepriivymasklihkqdervtgnggnlyfanvatsdnshdyic	209
Qy	241	---VGTNMGERESEVAELTILRSPVKRPSNIAVTVDSAE-----FKACE	286
Db	210	hahfgttrllqieq-ldlrvtasmidrkprlflptnsshlvalggpvlclciae	267
Qy	287	GDGPVTVKWRDGGELPKSRTEIRD-DHTLKIRKVTAGDMSGTVCVAZNMKGABASALT	345
Db	268	gftptklwlpssgmpdrvtqynhktlqlkvgeedgyerclaenslgasrharvy	327
Qy	346	TVQEPPEHVVRKQDVALGRVITFQCEATGNPQAPAIKWRREGSONLLFSYQPOSSSR	405
Db	328	twesapvhlkpsghlyggetatridcqvgrpwtwring-----ipwealkdgyk	382
Qy	406	SVSQGDLTILNVQRSDGYIICQLNWAGSITIKAYLEVTVDIADRPVPIRGFWNQT	465
Db	383	rl-grgallinsvqsdgtvqcearhlllanelylyqvql---pakiltad-nqt	435
Qy	466	-VAVDG-TFVLSCVATGSGVPIILA-RKDGVLSTQDSRIKOLENVGLQIYTKALGDTGR	522
Db	436	vnavostavtlckagaypsvsvledgtvlgdgerffwanotclairdaand	494

QY 523 YTCATSTPSGSEATWSAIEVQEFVGVPPPPPT-----DPNLISSP----- 562
 Db 495 yfclaaqdnqnmankvkdattgqstiekgkrvftocqsgvdpqlgpeitwr 554
 QY 563 -----APSKPEVTD----- 571
 Db 555 ggdgrldqldgskdyfiedgrlvihslidsqngscvasteldvresqllvsgpgp 614
 QY 572 -----VSRNTVLSQWPNLNSGATPTSYIIE-APFSHAGSSQWTAENVKETS 619
 Db 615 vprlvlsdhlitqsgvrvswspaedhnaplekyldefedkenapekyslgkvpngcts 674
 QY 620 AINGKAPNAILFLVRAANAYGISDPSISDPVKTQDVLTSQGVDRHKVQRELGNAYLH 679
 Db 675 tlklspvyhytfrvtalokygpgpsvsetvtpaapeknvdrkqegnetnmvi- 733
 QY 680 LHNPTVLSSSSIEVHWTVDOQSQVYIGQYKILRPSGANGHGESDNLVFEVTRPAKNSVYIP 739
 Db 734 -----twkplrwmadwnapqvg-yrvwrgpgt-----rpgwqeqivsdp-----flvrs 776
 QY 740 DLANGVNEIKARPPFNEFGQADEIKFAKLEAPSAAPP--QGVTVSKNDGNGTALVRS 797
 Db 777 atstfpyeikqavnsqgkpepytytysgedypgalpeiglell-----nssavik 832
 QY 798 WOPPPEDTONQWQVEYKV--WCLGNTRY--HINK-----TVDSSTFSVIFPLVPGIRTS 849
 Db 833 wrpdpkqghlrgsyovtywregsgqrkshkrhkhkhvvpantstvlslgrlpyssyh 892
 QY 850 YEVAASTGAGSGVSKSE-----POFIQDANGH-----PVS----- 879
 Db 893 lveqafogrgsgspaseftfstpegvpgpaealhecgscstslilrwppolshngltyvt 952
 QY 880 -----PEDVSLAQISDVVKQPAFIAGIACWAILMVFSLWYRHRKKRGLTSTYAGI 935
 Db 953 lsyldpdeqky-qslfnrdpel-----rthmltdlspgl 987
 QY 936 RVPSPFTPEPTVYTORQ-GEA-VSSGFRGLL-----NISEPAQVPLADTW-PWTC 984
 Db 988 r----rfrlqattkeqgealvregtmaigsldfignistagetyvswvpkq 1041

RESULT 14
 W42087
 ID W42087 standard; Protein; 1571 AA.
 XX
 AC W42087;
 28-SEP-1998 (first entry)
 DE Human Down syndrome-cell adhesion molecule DS-CAM2.
 DE
 KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
 KW signal transduction; trisomy 21; mental retardation;
 KW holoprosencephaly; corpus callosum agenesis;
 KW schizencephaly; diagnosis; assay; human.
 XX
 OS Homo sapiens.
 XX
 PN W09817795-AL.
 PD 30-APR-1998.
 XX
 PF 23-OCT-1997; 97W0-US19547.
 XX
 PR 25-OCT-1996; 96US-0029322.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Korenberg JR;
 XX
 DR WPI; 1998-21791/24.
 DR N-PSDB; V31988.
 XX

PT New isolated Down's Syndrome-cell adhesion molecule - used to
 PT develop products for detection, diagnosis and therapy of
 PT developmental and neurological abnormalities
 XX
 PS Claim 2; Page 90-95; 109pp; English.
 XX
 CC This polypeptide comprises Down syndrome-cell adhesion molecule
 CC DS-CAM2, an extracellular soluble protein belonging to a novel
 CC subclass of the Ig superfamily with highest homology to neural cell
 CC adhesion molecules. Its amino acid sequence was deduced from cDNA
 CC clones (see V31982) isolated from a trisomy 21 foetal brain library.
 CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and
 CC lacks the entire transmembrane domain of DS-CAM1. The invention
 CC provides human and murine DS-CAM nucleic acid sequences (see also
 CC V31981, V31985-87), expression vectors and host cells, transgenic
 CC animals, antibodies, antisense oligonucleotides, and primers
 CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
 CC with developmental and neurological processes. They can be used in
 CC e.g. neural prosthetic devices used in entubulation methods of
 CC repairing (regenerating) damaged or severed peripheral nerves, and
 CC also in bioassays to identify agonists and antagonists. The products
 CC can also be used in detection, diagnosis and therapy of developmental
 CC and neurological abnormalities such as Down syndrome, mental
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,
 CC or schizencephaly.
 XX
 SQ Sequence 1571 AA;

Query Match 8.1k; Score 708.5; DB 19; Length 1571;
 Best Local Similarity 26.8k; Pred. No. 3.2e-29;
 Matches 286; Conservative 147; Mismatches 446; Indels 187; Gaps 48;
 QY 64 EDPPFPIRVEHSDLVKSGEPALNCKAGRPPTTIEVWQGEVETOKDOP-----RSHR 119
 Db 403 edgtqklsafsekvsvpaepvslncnkvtpitltv-----ltdppllkgsgrh 454
 QY 120 -----MLPSSGLFFIRIVHGRSRDPDEYVYVCVARTVIGAVSHASLEVALRDFRQNP 176
 Db 455 isqmtelagrvslinssqvr-dggyvrtcnasag-vvlyglarlv-----rgpasirp 509
 QY 177 SDWVAV-GEPAWECOPPGRPPEPTISWKQSDPLDDKDRIT-ROGKLMTITRES- 233
 Db 510 mkaitelagrdtylhr-vigyypysikwyknsllpfhnrhgvafengtklisdvqkev 568
 QY 234 DAGHTVCGVMTWNGERE-----SEVAELTVLERPSFVRKPNLAVTVDSAEFXC-ZARGDP 289
 Db 569 degeytc-----vlyvgqlstsgsvhvtv-kypflgpfefgrfslggrflpovvsgdl 624
 QY 290 VPTVWRKKGDELKPSRYEIRD-----HTLIRAKTVAGMGSTTCVAVNMGKAEASATL 345
 Db 625 piltwkgdrgpipslygtionldfstsrlsnlslmhngnyticlarneaaavehqsll 684
 QY 346 TVQEPHPHFVPRQDQVAGTRVTFQCEATGNPQAPVWRGSGSNLFS-----TOP 398
 Db 685 lrvpvgkfvvrgdgdygkavilncsaesgypvptlvk-----fsgkqvgpvgp 736
 QY 399 PQSSSRFVSQVGTDLITNVORSDDGVYICQTLNAGSIIKTA-YLEVTVDIADPPPPV 457
 Db 737 lalngriqlvngsllkhvveedsgyckvndgadvksmyltv-----kipami 790
 QY 458 RQGVPTVTVAVDQ-TFVLSCAVTSQVPTVILNRKQDVLTSQDSR--IKQLENG-----V 509
 Db 791 tsyp-ntlaagggkemsctahgeklpvrwekdrlinpenarylvtvkegeevist 849
 QY 510 LQIRYAKLGDIGRYTCIASTPSGEATWSAIEVQEFVGVPPPPPTDNLIPSAPSKEPV 569
 Db 850 lqlplvredsgffschainsygedrglqltwe-----ppdppeliei 893
 QY 570 TVDSRNTVLSQWPNLNSGATPTSYIIEAFSHAGSSQWTAENVK-----TETSAIKGL 624
 Db 894 kvdkaitrltwmgfdngspitgydieckn-kdsdws-aqrtdkdvspqlnatiidi 950

Qy 625 KPAIYVFLVRAAAYGISDPSQISDPVKYTDVLPSTSQGVDEKQVRELGNVHLNRP 684
 Db 951 hpsstysirymaknrigksepsn-elitadeapdpqpg-----whe--- 994
 Qy 685 VLSSSSIEVHTVDQ---QSQYIQYKILYRPSGANHGSDNLVFEVTPAKNSV-VIPD 740
 Db 995 pssqsgirvkwapkhlgnglirgyqlyreystg-gnqfnlsvdtsqdsweyldn 1053
 Qy 741 LRKGVWEIKARPTFNEFGADSEIKFARTLEAPSPAGPOQTVSKNDNGTALVSWOP 800
 Db 1054 inkftgyvlvqaqractgtpssqelittledvpsyppevqaiat--speslsiswt 1111
 Qy 801 PPEDTQMGVQVETV---WCLGNETHYKINKYDSTFSVPIPLVPGIRFSEVVAATGA 858
 Db 1112 lskaelngllqgfrvlywanlmdgelgelnittqpsleldgkytnysiqvlfatra 1171
 Qy 859 GSGVKSEPOPIOL---DAHGNVSPEDQVSLAQGISDVVKQAFIAGIGAACWILLMVESI 916
 Db 1172 gdsrvseqlfrtkedvpgpp-----ag'kaaaasaswivs 1208
 Qy 917 WLVRHKKKNCILSTYACIRKVPSTPTPTVTVRGEAVSGGRGLNISEPAAPML 976
 Db 1209 wl--pklngi-----lryktvfcshpypt-----lsefasp-- 1241
 Qy 977 ADTW---PWTGNW---HNDCSISCTAGNNSDNLUT---YSRPDCIANTNQLDNQK 1027
 Db 1242 -dsfyrpnlrarnrgyvvwvavtsagrnsseilveplakaparilfsqvtvtpm 1300
 Qy 1028 TWMLPESTVIGVDLGNKINEMKTFN-SPNLK--DGR---FVNPVS 1067
 Db 1301 kdvlvpckav---gdsapavkwmkdsngtspvltdgrsifsnsg 1343

RESULT 15

W42086

ID W42086 standard; Protein; 1910 AA.

XX

AC W42086;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM1.

XX

KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;

KW signal transduction; trisomy 21; mental retardation;

KW holoprosencephaly; corpus callosum agenesis;

KW schizencephaly; diagnosis; assay; human.

XX

OS Homo sapiens.

XX

Key Location/Qualifiers

Peptide 1..23

FT /label= Sig_peptide

FT Protein 24..1910

FT /label= Mat_protein

FT Domain 24..887

FT /label= IG

FT /note= "Immunoglobulin type-C2 domain"

FT Domain 888..1594

FT /label= FcN

FT /note= "fibronectin type III domain"

FT Domain 1595..1616

FT /label= Transmembrane

FT Domain 1617..1910

FT /label= Cytoplasmic

FT Region 24..126

FT /label= Igl

FT Region 127..225

FT /label= Ig2

FT Region 226..316

FT /label= Ig3

FT Region 317..409

FT /label= Ig4

FT Region 410..506
 FT /label= Ig5
 FT Region 507..603
 FT /label= Ig6
 FT Region 604..697
 FT /label= Ig7
 FT Region 698..792
 FT /label= Ig8
 FT Region 793..887
 FT /label= Ig9
 FT Disulfide-bond 46..102
 FT Disulfide-bond 145..197
 FT Disulfide-bond 246..293
 FT Disulfide-bond 335..385
 FT Disulfide-bond 428..484
 FT Disulfide-bond 525..575
 FT Disulfide-bond 617..669
 FT Disulfide-bond 711..766
 FT Disulfide-bond 809..865
 FT Disulfide-bond 1307..1359
 FT Modified-site 78..80
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 106..108
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 470..472
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 487..489
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 558..560
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 666..668
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 710..712
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 748..750
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 795..797
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 924..926
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1142..1144
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1160..1162
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1250..1252
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1271..1273
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1324..1326
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1341..1343
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 1488..1490
 FT /note= "Asn is N-glycosylated"
 XX W09817795-AL.
 XX
 PD 30-APR-1998.
 XX
 PF 23-OCT-1997; 97WO-0519547.
 XX
 PR 25-OCT-1996; 96US-0029322.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Korenberg JR;
 XX
 DR WPI; 1998-271791/24.
 DR N-PSDR; V31981.
 XX
 PT New Isolated Down's Syndrome-cell adhesion molecule - used to
 PT develop products for detection, diagnosis and therapy of
 PT developmental and neurological abnormalities

[illegible]

Search completed: January 22, 2001, 12:19:37
Job time: 1734 sec

QY	64	EDFPFFPIVHEPSDLIVSGKEPALINCAKGRPTPTIIVWYKGGERVETDKDDP-----RSR	119
Db	403	edgtgpkliarsfekvcsaepvyslmcnvgtktpitvt-----tllddpllkgsgr	454
QY	120	-----MLPSSGFLFFRLVHVRKSRDEGVYCVVARNYLGEAVSNASLEVALLRDQRNP	176
Db	455	isqmtsegnvsvylnissqqv-dggyvrcntannsg-vlyqarinv-----rpsasir	509
QY	177	SDVVAV-GEPAWVKQCPGPRGPEPTTISWKKGSPLDDKERIT-RGKLMIITIKRS-	233
Db	510	mknatisagrtgythcr-vlgppyysskwyknsnlpfnhrqvfagnsgklksdvqev	568
QY	234	DAGKYVCVGVIMVHGER--SEVALAEVLIERFSVFKRPSMLAVTDSAEFK-ERAGDP	289
Db	569	degeytc-----arlvqqlgstssvhtvr-kvpplqgfefprisigrfvlpvcvsgsl	624
QY	290	VPTVWRKRDAGLEPKSRTEYRDD-----HTLKRIVKQDMSGTVAEWNKGAASATL	345
Db	625	plitcvgdgrppislgwtindifdssrlsrlslnhmqyqfclarcnaaevahesgl	684
QY	346	TVQEPHFHVFPKQVVALGRITVTCQCATGNPQPAIFWARGESQNLFS-----YOP	398
Db	685	lvrvpklvtpvqrqdglygkavlnscasgeyppvtvkw-----fskagvypgfp	736
QY	399	QSSSSRFSVSQTDGLTITWQSDGVYIQCITLVAGSIIITKA-FLVETDIADRPPTVI	457
Db	737	lalngriqlvngslilkhvrveesgrylcvsnvdgadvksmylt-----kipami	790
QY	458	RQGPVQIVTAVFD-VYGLSCVATGSPVPTILWRKQDVSITQDSR-ITQLENG-----Y	509
Db	791	tsyp-attlagagkkmctahgekpilvrveedrlinpenarylvstkevgeerist	849
QY	510	LQIRITAKLGDTRGTCTIAGTSPSGRATWSAIIIVQEGVPVQPPRPDNLNIPASPEV	569
Db	850	lqlptvredsgfscfhansygeedrgilgtwge-----ppdppeiei	893
QY	570	TDYSRTVTLSMOPNLNSGATPTIILAFBSAGSSGSMVAENVN-----TETSAIKGL	624
Db	894	kdvakrtitlwtmgfdgnsplitydieckn--ksdsws-aqrgtdvsgnslnatidi	950
QY	625	KPNAILYFLVRAANAGIDSPQISDVPKQDVLPTSGCVHQQVQREGLNVALLHNPT	684
Db	951	hpsstyslrwkrkrlsepsn-elitidaeagdpooe-----wile-----	994

Qy	1	MKKKVPFLWISLLSSPNELFLAQIPDPEDVGRGDNGRTPTTSDNDNDSLQVTGSR	60
Db	1	MKKKFLFLWISLLSSPKKKHLLAQIPDPEDLGRGNGRTPTTSDNDNDSLQVTGSR	60
Qy	61	LRQEDPPFPIVHPSDLVSKSGPATLNKCAAGRPPTIEWKGGKGVETKDDQASHM	120
Db	61	LRQEDPPFPIVHPSDLVSKSGPATLNKCAAGRPPTIEWKGGKGVETKDDQASHM	120
Qy	121	LLPSSSFLFLIVHGRSKSPDGGTVCAVNRIGBAVSHNASLEVALLRDQFQNSDYM	180
Db	121	LLPSSSFLFLIVHGRSKSPDGGTVCAVNRIGBAVSHNASLEVALLRDQFQNSDYM	180
Qy	181	VAYGEPAVMEQPPRGHPPTLSWKKGSPLLDDXERTIRGKKLMIITTKSDAGKTC	240
Db	181	VAYGEPAVMEQPPRGHPPTLSWKKGSPLLDDXERTIRGKKLMIITTKSDAGKTC	240
Qy	241	VGTMNGERGESVAELTVLERPSFKVRPNMLATVDSAEFKCAERGDVPVTVWKKDQ	300
Db	241	VGTMNGERGESVADYTVLERPSFKVRPNMLATVDSAEFKCAERGDVPVTVWKKDQ	300
Qy	301	ELPKSRTEIRDHLLKIRVITAGDMSGTCVAENMGKAEASALTVOEPHFHVKPRDQ	360
Db	301	ELPKSRTEIRDHLLKIRVITAGDMSGTCVAENMGKAEASALTVOEPHFHVKPRDQ	360
Qy	361	VVALGRTVFQCEATGNPAPLWRRGSGNLLFSYQPPGSSSFSVSQTDGLITNVR	420
Db	361	VVALGRTVFQCEATGNPAPLWRRGSGNLLFSYQPPGSSSFSVSQTDGLITNVR	420

Qy 421 SOWGYIQQTLNAGVSIITKAYLEVTDVIADRPVPVIRQGPVQTVADGTVFLSCVAG 480
Db 421 SOWGYIQQTLNAGVSIITKAYLEVTDVIADRPVPVIRQGPVQTVADGTVFLSCVAG 480
Qy 481 SPVPTILMRKQDVLVSTQDSRIKQLNGVLQIRYAKLGDGRTCTASTPSGEATWSAT 540
Db 481 SPVPTILMRKQDVLVSTQDSRIKQLNGVLQIRYAKLGDGRTCTASTPSGEATWSAT 540
Qy 541 EVQEGVVPVQPPPTDPLNIPSAKPEVTVDSKNTVLSWQNLNAGATPTSYIIAFA 600
Db 541 EVQEGVVPVQPPPTDPLNIPSAKPEVTVDSKNTVLSWQNLNAGATPTSYIIAFA 600
Qy 601 BASGSSQTVADNWKETSAIRGLKPNAILFLVRAANAYGISDPSQSDPVKTQVLP 660
Db 601 BASGSSQTVADNWKETSAIRGLKPNAILFLVRAANAYGISDPSQSDPVKTQVLP 660
Qy 661 SQGVDRVQVQRLGNVLAHNPVTLSSSSIVEMTVDDQSQIQGTILIRFSGASGE 720
Db 661 SQGVDRVQVQRLGNVLAHNPVTLSSSSIVEMTVDDQSQIQGTILIRFSGASGE 720
Qy 721 SGMVLEVPVTKNSVYIPDLRGVNYEIKARPVFNFGQADSEIKAFKLEAPGAPP 780
Db 721 SGMVLEVPVTKNSVYIPDLRGVNYEIKARPVFNFGQADSEIKAFKLEAPGAPP 780
Qy 781 GVTYSKNGGTAIVLSWQPPEDTQNGWQVYKWCGLNETRTHINKTVDGSTFVSVIP 840
Db 781 GVTYSKNGGTAIVLSWQPPEDTQNGWQVYKWCGLNETRTHINKTVDGSTFVSVIP 840
Qy 841 FLVPIGRSIVEVAAGSGGVKSEPPQFIQLDNGNVPSEDDVSLAQISDVKQPAFI 900
Db 841 FLVPIGRSIVEVAAGSGGVKSEPPQFIQLDNGNVPSEDDVSLAQISDVKQPAFI 900
Qy 901 AGTGAACWILMYSIKLYRHRKRNGLTSTYAGIRKVPSTPTPTVTVQGGVAVSSG 960
Db 901 AGTGAACWILMYSIKLYRHRKRNGLTSTYAGIRKVPSTPTPTVTVQGGVAVSSG 960
Qy 961 RPLGLNLESPAAQPLADTWPTNGNNHNSISCTAGNGNSDLNLTYSRPAACIANTY 1020
Db 961 RPLGLNLESPAAQPLADTWPTNGNNHNSISCTAGNGNSDLNLTYSRPAACIANTY 1020
Qy 1021 WQLONKQTHMLPESTVYGDVDSLKNINEMKTFSNPLKDRGFVPSQPTPIATTQLQ 1080
Db 1021 WQLONKQTHMLPESTVYGDVDSLKNINEMKTFSNPLKDRGFVPSQPTPIATTQLQ 1080
Qy 1081 SNLSNNMNGGSGSSEKHWPGLQKQVAPVQYNIYQNKLNKTRANDTVPTPIYQ 1140
Db 1081 SNLSNNMNGGSGSSEKHWPGLQKQVAPVQYNIYQNKLNKTRANDTVPTPIYQ 1140
Qy 1141 SYDQNTGSSYNSDRGSSYSGSQHKGARTPKVPKQGGNMADLPPPPHPPHNSSE 1200
Db 1141 SYDQNTGSSYNSDRGSSYSGSQHKGARTPKVPKQGGNMADLPPPPHPPHNSSE 1200
Qy 1201 EYNISVDESTQEMCPVPPARMYLQDELEEDERGTPTVPRGAASSPAAVTSQST 1260
Db 1201 EYNISVDESTQEMCPVPPARMYLQDELEEDERGTPTVPRGAASSPAAVTSQST 1260
Qy 1261 ATLTPSQBELQPMQLQCPETEGMHQPDRRQVPSVPPPPRPSIPHTGYISGLPV 1320
Db 1261 ATLTPSQBELQPMQLQCPEDLGMHPDPRRRQVPSVPPPPRPSIPHTGYISGLPV 1320
Qy 1321 DMDTAPPEEEDZADMEVAKQTRRLLLGLQETPSSVYGDLESSTVSGMNGWSAGE 1380
Db 1321 DMDTAPPEEEDZADMEVAKQTRRLLLGLQETPSSVYGDLESSTVSGMNGWSAGE 1380
Qy 1381 DNISGSRSSVSSGSGFTTADFAQAVAAAYAGLVARRQMQAAGRHFASQPRP 1440
Db 1381 DNISGSRSSVSSGSGFTTADFAQAVAAAYAGLVARRQMQAAGRHFASQPRP 1440
Qy 1441 TSPVSTDSMSAAVQTRPAKLLKHQPHLRRETYDOLPPPPVPAIKSPQAQSTQ 1500
Db 1441 TSPVSTDSMSAAVQTRPAKLLKHQPHLRRETYDOLPPPPVPAIKSPQAQSTQ 1500
Qy 1501 LEVPRVVPVVKLSMDARTDRSSDRGSSYKREVLQGVVMNRNPGDPRQAQDNG 1560

Db 1501 LEARPMGPKLASIEARDRSSDRGSSYKREVLQGVVMNRNPGDPRQAQDNG 1560

Qy 1561 KGRGNAAKREDLPKAKTHLQEDILPYCRPTPTSPNNPDPSSSSSMSSRSGSRQEDA 1620

Db 1561 KARGTAKREDLPKAKTHLQEDILPYCRPTPTSPNNPDPSSSSSMSSRSGSRQEDA 1620

Qy 1621 WGRNMAEMVGLGFERGDNNELEETES 1651

Db 1621 WGRNMAEMVGLGFERGDNNELEETES 1651

RESULT 2

T30805

dutt1 protein - mouse

N:Alternate names: transmembrane receptor protein Robol homolog

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 | sequence_revision 22-Oct-1999 | text_change 22-Oct-1999

C:Accession: T30805

R.Wu, M.C.; Lowe, W.; Fordham, R.; Rabbitts, P.

submitted to the EMBL Data Library, July 1998

A:Description: The mouse homologue of human DUT1/H-robol gene: protein sequence and

A:Reference number: z20879

A:Accession: T30805

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1612 <W>

A:Cross-references: EMBL:117793; NID:el329712; PID:el329713; PIDN:CAA76850.1

A:Experimental source: brain

C:Genetics:

A:Gene: dutt1

A:Map position: 16

Query Match 93.14; Score 8120; DB 2; Length 1612;
Best Local Similarity 95.64; Pred. No. 0;
Matches 1525; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

Qy 57 TGSRLAQDPPFRIYERHPSDLISYKGPATLNKAGRPPTIIEWYKGERVETDKDOP 116

Db 18 SGRSLAQDPPFRIYERHPSDLISYKGPATLNKAGRPPTIIEWYKGERVETDKDOP 77

Qy 117 SHKMLPSSLIFFLRIVHGRKSRPDGVTYCVARNLGEAVSHNASLEVALRDQFRNP 176

Db 78 SHKMLPSSLIFFLRIVHGRKSRPDGVTYCVARNLGEAVSHNASLEVALRDQFRNP 137

Qy 177 SDVMVAGEPAMVQCPRGHPPTISNKKDGPLDOKDERITIIRGKLMITITRKSADG 236

Db 138 SDVMVAGEPAMVQCPRGHPPTISNKKDGPLDOKDERITIIRGKLMITITRKSADG 197

Qy 237 KYVCVGNWGERSEVAELTVLERPSVKRPSNLAVTYDQSAEFCARGDPVPTVNR 296

Db 198 KYVCVGNWGERSEVAELTVLERPSVKRPSNLAVTYDQSAEFCARGDPVPTVNR 257

Qy 297 KDQGLPKSRTEIRDDTLKIRKVTAGDMGSTYCVARNWKAASATLVQEPHFHYVK 356

Db 258 KDQGLPKSRTEIRDDTLKIRKVTAGDMGSTYCVARNWKAASATLVQEPHFHYVK 317

Qy 357 PRDQVWALGRTVTFQCEATGNPQPAIFWRRGSQLFLSTYQPOSSSRFSVSTQGLTI 416

Db 318 PRDQVWALGRTVTFQCEATGNPQPAIFWRRGSQLFLSTYQPOSSSRFSVSTQGLTI 377

Qy 417 NVQRSOWGYIQQTLNAGVSIITKAYLEVTDVIADRPVPVIRQGPVQTVADGTVFLSC 476

Db 378 NVQRSOWGYIQQTLNAGVSIITKAYLEVTDVIADRPVPVIRQGPVQTVADGTVFLSC 437

Qy 477 VATGSPVPTILMRKQDVLVSTQDSRIKQLNGVLQIRYAKLGDGRTCTASTPSGEATW 536

Db 438 VATGSPVPTILMRKQDVLVSTQDSRIKQLNGVLQIRYAKLGDGRTCTASTPSGEATW 497

Qy 537 SAYIEVQEGVVPVQPPPTDPLNIPSAKPEVTVDSKNTVLSWQNLNAGATPTSYII 596

Db 498 SAYIEVQEGVVPVQPPPTDPLNIPSAKPEVTVDSKNTVLSWQNLNAGATPTSYII 557

Qy 597 EAFSHAGSSGWSQVAENWKTTSTAIGLKNPAIYFLVRAANAYGISDPSQISDPVKTD 656
 Db 558 EAFSHAGSSGWSQVAENWKTTSTAIGLKNPAIYFLVRAANAYGISDPSQISDPVKTD 617

Qy 657 VLPTSGVDKQVORELGNVLAHLNPTVLSSSIEVHWTVQQSQYIQTKLYRPSGA 716
 Db 618 VPTSGVDKQVORELGNVLAHLNPTVLSSSIEVHWTVQQSQYIQTKLYRPSGA 677

Qy 717 NHGSDMLVFVETPAKNSVVPDLRKGNYEIKARPFFNEFGADSEIKFAKLEAPAS 776
 Db 678 SHGESMLVFVETPTKNSVVPDLRKGNYEIKARPFFNEFGADSEIKFAKLEAPAS 737

Qy 777 APPQGVTSKNDNGTALLVSWQPPEDTQNGMVQYKVMCLNGETRYHINKTVDSGTS 836
 Db 738 APPRSVTSKNDNGTALLVSWQPPEDTQNGMVQYKVMCLNGETRYHINKTVDSGTS 797

Qy 837 VYIPLVPIGRTSYEVAASGAGSGYKSEPOIQLDQAHNPVSPEDQSLAQOISDVVK 896
 Db 798 VYIPLVPIGRTSYEVAASGAGSGYKSEPOIQLDQAHNPVSPEDQSLAQOISDVVK 857

Qy 897 PAFIAGIACAIIWLWFSILYRHRKRNGLSTYAGIRKVPSTPTPTTYTQRGGEAV 956
 Db 858 PAFIAGIACAIIWLWFSILYRHRKRNGLSTYAGIRKVPSTPTPTTYTQRGGEAV 917

Qy 957 SSGRGGLNISEPAAQFWLADTWPNNGNNDIS13CCTAGNGNSNLNTTSPADCI 1016
 Db 918 SSGRGGLNISEPAAQFWLADTWPNNGNNDIS13CCTAGNGNSNLNTTSPADCI 977

Qy 1017 ANYNKLNQKQNLMLPESTYGVDLSENKINEMKTNSPNLKDGRVNPSSGQPTTYAT 1076
 Db 978 ANYNKLNQKQNLMLPESTYGVDLSENKINEMKTNSPNLKDGRVNPSSGQPTTYAT 1037

Qy 1077 QLQISLNNMNGSGDSGSKHNPQOQGVAPVYNQNLKMDYRANTVPPPTI 1136
 Db 1038 QLQISLNNMNGSGDSGSKHNPQOQGVAPVYNQNLKMDYRANTVPPPTI 1097

Qy 1137 PYNGSTQMGVSYNNSDGSSTSGSGHKKARTKPVKQCGMMWDLPPFPAPPH 1196
 Db 1098 PYNGSTQMGVSYNNSDGSSTSGSGHKKARTKPVKQCGMMWDLPPFPAPPH 1157

Qy 1197 SNSEYINISVDSYDQEMPCVPAARMYQLQDELEEDERGPTPPVGAASSPAANVTS 1256
 Db 1158 SNSEYINISVDSYDQEMPCVPAARMYQLQDELEEDERGPTPPVGAASSPAANVTS 1217

Qy 1257 HQSTATILTPSQELQPMQLQCEPETHQMQDORRQPVSPPPPRTISPPHTTGIS 1316
 Db 1218 HQSTATILTPSQELQPMQLQCEPETHQMQDORRQPVSPPPPRTISPPHTTGIS 1277

Qy 1317 PLVMSMDTAPAEEDDAEAMVAKMTKRLLRGLTGPASSVGLDESSYTGSMINGWS 1376
 Db 1278 PLVMSMDTAPAEEDDAEAMVAKMTKRLLRGLTGPASSVGLDESSYTGSMINGWS 1337

Qy 1377 ASEEDNISGSGSVSSGDSFFTDADPAQAAAEAGLKVARRQMQDAAGRRHFASQ 1436
 Db 1338 ASEEDNISGSGSVSSGDSFFTDADPAQAAAEAGLKVARRQMQDAAGRRHFASQ 1397

Qy 1437 CPBPTSPVSTDSNMAAVMKTBPAAKLKQPHLRRETTIDCLPPPPVPAIKSPTAQ 1496
 Db 1398 CPBPTSPVSTDSNMAAVMKTBPAAKLKQPHLRRETTIDCLPPPPVPAIKSPTAQ 1457

Qy 1497 SKYQLVFPVYVVKPLSMDARTDSRSGSYKRGVLDGRQVDMRTNPGDPRAEQ 1556
 Db 1458 SKYQLVFPVYVVKPLSMDARTDSRSGSYKRGVLDGRQVDMRTNPGDPRAEQ 1517

Qy 1557 QNDGCGNNAKAKRDLPAKTHLQEDILPCRCPTPTTNNPDRSSSSMSRSGSGSR 1616
 Db 1518 QNDGCGNNAKAKRDLPAKTHLQEDILPCRCPTPTTNNPDRSSSSMSRSGSGSR 1577

Qy 1617 RQDANVGRNIAQMVLGGFERGDENNELEETES 1651
 Db 1578 RQDANVGRNIAQMVLGGFERGDENNELEETES 1612

RESULT 3

T14316

rig-1 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 20-Sep-1999.#sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14316

R:Yuan, S.S.F.; Cox, L.A.; Dasika, G.K.; Lee, E.Y.H.P.

submitted to the EMBL Data Library, April 1998

A:Reference number: 217975

A:Accession: T14316

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1344 <TUA>

A:Cross-references: EMBL:AF060570; NID:g4206385; PID:g4206386; PIDN:AA011628.1

Query Match 29.9%; Score 2607.5; Db 2; Length 1344;
 Best Local Similarity 38.2%; Pred. No. 3.2e-115;
 Matches 594; Conservative 204; Mismatches 456; Indels 299; Gaps 31;

Qy 58 GSRLLQDEPPRIVHEPDSILVSKGEPATLNCRAEGRPTIEMWKGEERVDTDODPSS 117

Db 32 GSRVGPEDAMFVYQPPDLVSVSGEPATLPCRAEGRPTIEMWKGEERVDTDODPSS 91

Qy 118 HMLLPSGLFPPLIVHGRKSPDGVVYVARNVLEAGSHNASLEVALADQDPQWPS 177

Db 92 HMLLPSGLFPPLIVHGRKSPDGVVYVARNVLEAGSHNASLEVALADQDPQWPS 151

Qy 178 DWVWAGEPAVMEQPPRGHPTIISWKDGSPLDOKDERITISGGKMLITTRKSDAG 237

Db 152 DWVWAGEPAVMEQPPRGHPTIISWKDGSPLDOKDERITISGGKMLITTRKSDAG 211

Qy 238 TWCVGNMNGERESEVALVLERSPVKRPSLAVTVDSAEKCAEAGDPVPTVWRK 297

Db 212 TWCVGNMNGERESEVALVLERSPVKRPSLAVTVDSAEKCAEAGDPVPTVWRK 271

Qy 298 DQGLPKSRTIREDHLLKIRKVTAGDGSYTCVAENWKGAEASATLVYQPPFVYK 357

Db 272 DQGLPKSRTIREDHLLKIRKVTAGDGSYTCVAENWKGAEASATLVYQPPFVYK 331

Qy 358 RQVVALGRTVTCQATGNQPAIFVWRREGSNLLSPQSSSSFSVSQDGLTIN 417

Db 332 RQVVALGRTVTCQATGNQPAIFVWRREGSNLLSPQSSSSFSVSQDGLTIN 391

Qy 418 VQRSDWYITQTLNAGSIIKAYLEVDIADRPVPIRQVFNQTVAVDGTFLVSC 477

Db 392 VQRSDWYITQTLNAGSIIKAYLEVDIADRPVPIRQVFNQTVAVDGTFLVSC 451

Qy 478 ATGSPVPTILAKMDGLVSTQDSRIKQENGLVQIQTAKLGTGRTICTASPGEATNS 537

Db 452 ATGSPVPTILAKMDGLVSTQDSRIKQENGLVQIQTAKLGTGRTICTASPGEATNS 511

Qy 538 AYIEVQE-RGVVQPPRPTDPLNAPSAPKPEVDIVSRNTVLSWQPNLNGATPTYSI 596

Db 512 AYIEVQE-RGVVQPPRPTDPLNAPSAPKPEVDIVSRNTVLSWQPNLNGATPTYSI 569

Qy 597 EAFSHAGSSGWSQVAENWKTTSTAIGLKNPAIYFLVRAANAYGISDPSQISDPVKTD 656

Db 570 EAFSHAGSSGWSQVAENWKTTSTAIGLKNPAIYFLVRAANAYGISDPSQISDPVKTD 629

Qy 657 VLPTSGVDKQVORELGNVLAHLNPTVLSSSIEVHWTVQQSQYIQTKLYRPSGA 716

Db 630 VLPTSGVDKQVORELGNVLAHLNPTVLSSSIEVHWTVQQSQYIQTKLYRPSGA 689

Qy 717 NHGSDMLVFVETPAKNSVVPDLRKGNYEIKARPFFNEFGADSEIKFAKLEAPAS 776

Db 690 NHGSDMLVFVETPTKNSVVPDLRKGNYEIKARPFFNEFGADSEIKFAKLEAPAS 747

Qy 777 APPQGVTSKNDNGTALLVSWQPPEDTQNGMVQYKVMCLNGETRYHINKTVDSGTS 836

Db 748 APPQGVTSKNDNGTALLVSWQPPEDTQNGMVQYKVMCLNGETRYHINKTVDSGTS 807

Qy 837 VVIFVLPGIRISVEVAASGAGSVKSEPOFIQL--DAHGNVSPEDQVSLAQISDVV 694
 Db 808 VTFSGLLGPQITRALVAATASAGVVASAPVLQVLPFAEP-GPEVSEGLAEKALV 866
 Qy 895 KQPAFIAGIAGAACMLMVFSIWLRYHKKRKLSTYAGIRKVPSTFTPTVYORGE 954
 Db 867 KKPAPLGSAAAGCALLGPCAALYRQKQKELSHYTA-----SFATPAVSPHSG 920
 Qy 955 AVSSGGGP--GLNLTSEPAQWADWTPTGNHNN--DCSISCTAGNCSNLTYS 1010
 Db 921 LGSLSRRPQGL--GPAAYPLADSWPHPPSPASQEPKSCCPS--NFD----- 966
 Qy 1011 RPADCIANYNQLDNQTNLMPESTVGDVLSKNINDEMTFNSPLKDGFRVNSGP 1070
 Db 967 -----PDRYTNAGISLYL-----AQTAGNASSG 994
 Qy 1071 TPTATITQLISLNNMNGSDSGEKKHKLQKQVAPVQIVIVQNKLMQFRAND 1130
 Db 995 FVTISD-----PVGEELQ----- 1008
 Qy 1131 TVPTPIPNQSDQWGGSSYSDSGSSGSGQGHKAGARTPKVPHQGGMWADLPPP 1190
 Db 1009 THGGFGPQRSSQDSTWSQITAPPENSGDSGARG-GQGLKLAGVQWPLSLWPEALPPP 1067
 Qy 1191 AHPPHNSSEETINVSDETDQMPCPVPAARMYQDELEEDERGPTTPVR----- 1244
 Db 1068 P-----SCELSCPBGP-----EELKGSOLEEWCVPVPEKSHLV 1102
 Qy 1245 GAASSPAV-----SYSHQSTATLTSPQSEQLQMLQCPETBGRMQHQPDR 1291
 Db 1103 GSSSGCYNAPAPROTPTSPSTSYGQSTATLTTPSPDPQP-----PTDIPHLQMP----- 1155
 Qy 1292 RRQVSPPPPPPTSPPTHTYISGLVSDMTDAPEEBEDADMEVAKMQRALLGL 1351
 Db 1156 KRVLPSPSP----- 1165
 Qy 1352 BQTPASVGDLESSTVGMINGNGSASEDNISGRSSVSDGSFPTDAQAVAAAA 1411
 Db 1166 -----LSVQPALSHRG----- 1178
 Qy 1412 ETAGLKYARRQMADAAGR-RHFHASQCPRTSPVSTDSNANVMQKTR-----AK 1462
 Db 1179 -----RPVGLGAGVLSYHASPSPVSTASAPGRTOVTGEMTPPLHGHRARIK 1229
 Qy 1463 KLKQKPHRLHRETYTDLPPPPVPAIKSPQAQSKTLEVPVVPVPLKMDARTDSS 1522
 Db 1230 KXPALP--YRREHSPGDLPPPLPPELADKLALGSA--GSRQVTPR-----ARAQWE 1280
 Qy 1523 DR-KGSSYHGREVLDRGVDMRTNPGPRAQSQDNGKRGKNAKRDLP 1574
 Db 1281 EBSGASASRG-----PTSSQGRPHDGKQSG-----GRRGLACRSFNSP 1321

RESULT 4

T42405

sax-3 protein - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000

C:Accession: T42405

R:Zellen, J.A.; Yi, B.A.; Bargmann, C.I.

Cell 92, 217-227, 1998

A:Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple asp

A:Reference number: Z22160; MUID:98117250

A:Accession: T42405

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1273 <L>

A:Cross-references: EMBL:AF041053; NID:g2804779; PID:AA38948.1; PID:g2804780

C:Genetics:

A:Note: sax-3

C:Function:

A:Description: sax-3 function is required at the time of axon guidance

Query Match 17.4%; Score 1515; Db 2; Length 1273;
 Best Local Similarity 33.5%; Pred. No. 8.2e-64;
 Matches 401; Conservative 164; Mismatches 469; Indels 164; Gaps 34;

Qy 68 PRIVEHPSDLVSKGEPATLNCAGRPTITWYKGGERVETDQDPSRHMPLPSGL 127
 Db 31 VTIHPRIIDVVSRSRGPATLNCAGK-PTAKITWYKGGVPTINKEQVSHRVLDTGSL 89
 Qy 128 PFILVYHGRKSR-PDGVYTVYANVLEAVSHNALEVALRDQKPNSDVMWAVEGP 186
 Db 90 PLLKRVKSNKWDAGATYCVASNEHGVSEHSGSLKLMREDFVRVPTVQALGSEM 149
 Qy 187 AVNCGPPHGPPEPTISWKDGSPLDDK-ERITIRG-GLMTITTKRSQAGTVCGVTN 244
 Db 150 AVLECSPPGTPPEPVSWKDKELAIQDMPTITLSDGNLIDPWSDSGTCQVANN 209
 Qy 245 MWGERSVEALTVLERPSFKVRSLNATVDSAEFKCARGDPVTVVHMKDGLPK 304
 Db 210 MWGVRSNPRLSVFEPKPEQKPMQVDGVAALFDCVPTGDPQQTITWKKNKMPV 269
 Qy 305 SR-YEIRDHDTLIRKIVTAGMSYTVARMWKAASATLVQPHFVWPKRQVWA 363
 Db 270 TRATKADMGRLIERYQPSDGETVCLARPGATLEASHLVAVFPSTQITADQVNP 329
 Qy 364 LGRTVITQCAAGNQPAITWRREGSNLLE-SYQPPSSSFVSQVGLTITVQASD 422
 Db 330 AGGTATFECTLQGPSPATWSKEQQDLFPST--VSADGRKTVSPITGLTIEVQVD 387
 Qy 423 WGYVICQLNAGSITIKALEVDIVAD-----RPPPIRGVQWQTVANDGTVFLSCV 477
 Db 386 SGATVCGAGNSGSSLSAALKYITKAVGTNPKPPTTIEGHQVNLVWGSAAFLPCV 447
 Qy 478 ATGSPPTILWAKDGLVNSTQSRILKQNGVQIYKALQDGTGTCIATSPGENTWS 537
 Db 448 ASGRPTPLGSHLRDGLIDITDSIRSQBSTSGHIALAKLPQGTCTYTIKANGDSWS 507
 Qy 538 ATYEVGFVGPVPPRTDNLPLSAPKSPETVDSRNTVLSWQNLNSGATP-TSTII 596
 Db 508 ASLTVEDHNAQVFMPPDSNFPSTPQITVIVWITDTEVLEHWNAPSTSGAGPTGT 567
 Qy 597 EAPSHSGSSWQTVANVKTETSAIKGLKNATLYFLVRAANAYISDQPSIDPVKT-- 654
 Db 568 QYTPDGLQWTFNIPDVTASTETRLKLPESHSMFVIRAEKNIGTQPSVSSALVTSK 627
 Qy 655 ----QDVLPTSGVDHKQVOREL-GRNVLHLNPTLVSSSIEVHWTDQSQSYIQGTAIL 710
 Db 628 PAQVQALSKRIKMDMAIAEKLTSGLKLEBVTINSTAVRLFKWKKRLLEEDITGIK 687
 Qy 711 YR-PSGANGESDVLVFEVTRPAKNSVIVPOLKAGVNYEIKARFP--FNEFGADSEK 766
 Db 688 WRGPPRTNQW---YVNTSPSTENYVSNLPMTHYEFYIPTHSGVSGISGAPSNM 743
 Qy 767 FANTLEAPSPAGQVTSKNDGNTALVSNQPPEDTQNGWQYKVMCLNETHRI 826
 Db 744 DVLTAAPPLPPEDVRIKL--NLTLIRISWAKPADGINSILKGPQITVIGQAPNWR 801
 Qy 827 NKTVDSTISVFIPLPGIRISVEVAASGAGSVKSEPOFIQLDAHGNVSPEDQVSL 886
 Db 802 NITNTERASTYLEHLTGTMYKIRVARSNGGV-----SGTSEVIMQDQL 851
 Qy 887 AQQISDVVQPAFIAGIACW-----LMVFSIWLRYHKKRKLSTYAGIRKVP 939
 Db 852 EKHLAAQQENESFLYGLINKSHVPPVIVALLIFVLIAYCTWRNSRSD--GDOR 907
 Qy 940 SFTPTPTVYORGEAVSSGVRGGLNLN-PAEQWADWTPTGNHNNDCSISCTAG 998
 Db 908 SF-----IKINGSVHMSN--NLDWVAQNPNQNPYNTAGRMNNRNGALSLPN 959
 Qy 999 NGNSDSNLTYS-----RPADCIANYNQLDNQTNLMPESTVGDVLSKNINDEMTFNS 1054
 Db 960 AQDFPNKDDISGTHMRPGSEHYHYAQLTGFGPNAM--STFTG-----NQYHD----- 1006

Qy 1055 SPMKDKGRFVNPSPQPTTYATTQLISLNNMNGSGDSGEKHVKPLGQKQVAVQY 1114
 Db 1007 -----DPSYATTTLWLSN-----QQ-----PAWL 1026
 Qy 1115 NIVEQNKMLKIDTRANDVPPPIPNQSVQNTQSGVSSSDGSS-----TSGS 1162
 Db 1027 N---DKMLRAPAMPNTNPPVPEPP--ARYADHTAGRRSSSRASDGGTLMGLHRTSGS 1081
 Qy 1163 Q-----GHHK-----GARTPKVPKQGMNWADLLPPFPAPPPP 1195
 Db 1082 QRSDDPHTDVTSLVQSLDGGSGSKERTGERTP--PKMTLM--DFIPPPSPNPP 1134

RESULT 5
 T29549
 hypothetical protein ZK377.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 Accession: T29549
 Nhan, M.; Hawkins, J.
 submitted to the EMBL Data Library, February 1997
 A:Description: the sequence of C. elegans cosmid ZK377.
 A:Reference number: Z20639
 A:Accession: T29549
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-423 (CDS)
 A:Cross-references: EMBL:U88183; PIDN:AAB52658.1; GSPDB:GWO0028; CESP:ZK377.3
 A:Experimental source: strain Bristol N2; clone ZK377
 C:Genetics:
 A:Gene: CESP:ZK377.3
 A:Map position: X
 A:Introns: 24/1; 142/3; 229/3; 284/2; 408/3

Query Match 9.84; Score 852; DB 2; Length 423;
 Best Local Similarity 46.18; Pred. No. 3e-33;
 Matches 177; Conservative 55; Mismatches 144; Indels 8; Gaps 7;

Qy 68 PRIVEPSPDLVSKGEPAFLNCKAAGRPPTIIEYKGGHVERVETQDPPSHRWLPSGSL 127
 Db 30 PVTHIEPDIWVSRSGPAFLNCKAG--PSTAKITWKDGPVITNKQVSHRVLDTGSL 88
 Qy 128 FFLIRIVHGKSR-PDEGVYCVARNYLEAVSNASLEVALLDQFQNPVDWVAVGEP 186
 Db 89 FLLAVKNGSGKSDGMAITCVASNEHGEVSNESGLKAMLRDEPKYRPTVQALGDEM 148
 187 AWMEQPPHGGPPTISHWKDGSLDQD--ERITING--GKLMITTRKSDAGKYVGVGN 244
 Db 149 AVLCESPPHGGPPTVYVSWKDDKELRIQDWRITLSDGNDLIDVDGSDGCTQVANN 208
 Qy 245 MYGERSVEAEVLTLERPSVFKRPSNLAIVDDSAEFCRAGDPVPTVWRKADGELPK 304
 Db 209 MYGERSVNRARLSVFEKPKFQEPQKMTVDVGAVLDFQVRYGDPQPIQTKWKNMEPV 268
 Qy 305 SR-TEIRDHDTLKIRKVTAGMSGSTVCAENMVKABASATLTVQEPHFVPRDQVVA 363
 Db 269 TRATYAKDNGKRIERVQSDSGEYTVARNPAGTLEASHLKRVAPPSFTQITADQVSP 328
 Qy 364 LGRVTITQCCATGNQPAIFWRREGSNLLP-SYQPPSSSRSFVSQDGLTITNQVRS 422
 Db 329 AGGTATFECTLVQGPSFAYFWSKEQDQLLEPSY--VSADGRKTVSPDTGLTIEVQVD 386
 Qy 423 WGYITQIVLNAGSIIITKYLEVT 446
 Db 387 EGATVCAKMSAGSSLSKALKAT 410

RESULT 6
 S05479
 neural cell adhesion molecule L1 precursor - mouse
 C:Species: Mus musculus (house mouse)
 Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S05479; B06050; S22167
 R:Moos, M.; Tacke, R.; Scherer, H.; Teplow, D.; Frueh, K.; Schachner, M.
 Nature 334, 701-703, 1988
 A:Title: Neural adhesion molecule L1 as a member of the immunoglobulin superfamily w
 A:Reference number: S05479; MUID:8818924
 A:Accession: S05479
 A:Molecule type: mRNA
 A:Residues: 1-1260 (CDS)
 A:Cross-references: EMBL:L12875; NID:G53336; PIDN:CAA31368.1; PID:G53337
 A:Note: the authors translated the coding CCT for residue 166 as Leu, ACT for resi
 A:Note: part of this sequence, including the amino end of the mature protein, was
 R:Rathjen, F.G.; Wolff, J.M.; Frank, R.; Bonhoeffer, F.; Rutishauser, U.
 J. Cell Biol. 104, 343-353, 1987
 A:Title: Membrane glycoproteins involved in neurite fasciculation.
 A:Reference number: A60850; MUID:87109457
 A:Accession: B60850
 A:Molecule type: protein
 A:Residues: 20-28, 'XX', 31-36 (CDS)
 R:Kohl, A.; Giese, K.P.; Mohajerli, M.H.; Montag, D.; Moos, M.; Schachner, M.
 submitted to the EMBL Data Library, December 1991
 A:Description: Analysis of promoter activity and 5' genomic structure of the neural c
 A:Reference number: S22167
 A:Accession: S22167
 A:Molecule type: DNA
 A:Residues: 1-165, 'L', 167-189, 'E', 191-281, 'S', 283-395, 'S', 397-514, 'APEKNPVDV', 524, 'GL
 A:Cross-references: EMBL:X63511
 C:Genetics:
 A:Introns: 26/1; 31/1; 66/2; 133/1; 174/1; 231/1; 268/2; 330/1; 374/1; 422/1; 452/2
 A:Note: The list of introns may be incomplete
 C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homolog
 C:Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmem
 F:1-19/Domain: signal sequence predicted <SIGD
 F:20-1260/Product: neural cell adhesion molecule #status experimental (CDS)
 F:256-313/Domain: immunoglobulin homology (CDS)
 F:440-498/Domain: immunoglobulin homology (CDS)
 F:531-592/Domain: immunoglobulin homology (CDS)

Query Match 8.84; Score 764.5; DB 1; Length 1260;
 Best Local Similarity 23.34; Pred. No. 1.8e-28;
 Matches 307; Conservative 177; Mismatches 510; Indels 325; Gaps 51;

Qy 56 YTSRLQDEDPFPAIVH--PSDLVSKGEPAFLNCKAAGRPPTIIEYKGGHVERVETD 112
 Db 26 YGSHVLE---PPVITEQSPRALVPPDIDSLKCEARGQVPTVWTKDGLHFAKPEEL 82
 Qy 113 ----DQPSRHMLPSGSLFLRLVHGKSRPDEGVYCVARNYLEAVSNASLEVAL 167
 Db 83 GVYVHAPTSQFSTIEGNSFARF-----QGITCTASNLKGTAMSH---ETQL 129
 Qy 168 LDDQFQNPQS-----VMVAVGEPVMEQCPGRHPEPTISNKKDGSPLD--KDRIT 220
 Db 130 VADGAKPKKPTVFKVPEVEGSESVLPCNPFPSPAAPPIYWM--NSKIPIDQGVSRVNG 187
 Qy 221 RGGKLMITTRKSD--AGKYVGVGNMGER-----ESEVAELTLERPSVFKRPSNLA 276
 Db 188 QNDGLIFANVLTSDNHSYDTC--NAPPGTIRILIKQEPDLRVKPTNMDIKRPLLEPT 246
 Qy 277 DSAB-----FKCEARGDPVPTVWRKADGELPKSRYEIRDDH--TLAKIRVTAG 324
 Db 247 SSSRLVALQGSLLIECTAGFPPTPIIKWLPSPDMPITRY--IQNKNTQLLNLVGEED 305
 Qy 325 MSGSTVCAENMVKABASATLTVQEPHFVYKRDQVVALGRTITQCCATGNQPAIFW 384
 Db 306 DGEYTCIAENSLGSAHAYTVVEAAPTLQKPSGLNPGSTARLQDQVGRQPEPTW 365
 Qy 385 RREGSNLLPSTQPPQSSSRSFVSQDGLTITNQVSDGYIYQIVLNAGSIIITKYLE 444
 Db 366 RING-----MSMETVNDQYRIEQ--GSLISLVQPTDMVQCEARNGHLLANAYIT 419
 Qy 445 VIDVIRADPPPIVIRQGPVNOT--VAVDG--FVLSCVATGSPVPTIWRKDGVLVTSQDS 502
 Db 420 VVOL-----PARILYKD--NOTVAVRGSTAYLCKAGFAPVSVQWLDREGLTVQDER 473

```

RESULT      7
A:1060
neural cell adhesion molecule L1 precursor - human
N:Alternate names: L1CAM
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A41060; S18454; A35331; S21971; S21972; A50223; A31072; G20506
R:Hlavin, M.L.; Lemmon, V.
Genomics 11, 416-423, 1991
A:Title: Molecular structure and functional testing of human L1CAM: an integrin
A:Reference number: A41060; MUID:92120663
A:Accession: A41060
A:Molecule type: mRNA
A:Residues: 1-1257 <HLA>
A:Cross-references: GB:M64296; NID:q186053; PIDN:ACL4352.1; PID:q3068548

```

Kobayashi, M.; Miura, M.; Asou, R.; Uyemura, K.
Biochim. Biophys. Acta 1050, 238-240, 1991
A:Title: Molecular cloning of cell adhesion molecule L1 from human nervous tissue: a
A:Reference number: B18454; MUID:92031698
A:Accession: B18454 :
A:Molecule type: mRNA
A:Residues: 1-3; 'V',5-215,'I',217-249,'T',251-275,'SV',278-356,'E',358-625,'V',627-12
A:Cross-references: DMBL:X59847; NID:g35009; PIDN:CAA42508.1; PID:g35010
A:Note: the authors translated the codon GAA for residue 27 as Gly
R:Jabali, M.; Wattel, M.C.; Nguyen, C.; Roux, D.; Demengeot, J.; Denizot, F.; Moos,
Genomics 7, 587-593, 1990
A:Title: the gene encoding L1, a neural adhesion molecule of the immunoglobulin fami
A:Reference number: A35331; MUID:90353957
A:Accession: A35331 :
A:Molecule type: DNA
A:Residues: 332-371 <DTA>
A:Cross-references: GB:MS5271
R:Rosenthal, A.; MacKinnon, R.N.; Jones, D.S.C.
Nucleic Acids Res. 19, 5395-5401, 1991
A:Title: PCR walking from microdissection clone M54 identifies three exons for <T
A:Reference number: S21971; MUID:92020233
A:Accession: S21971 :
A:Molecule type: DNA
A:Residues: 1082-1176 <ROS>
A:Cross-references: DMBL:X58775; NID:g29642; PIDN:CAA41576.1; PID:g29643
A:Accession: S21972 :
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 353-935,'V',937-1176 <RO2>
A:Cross-references: DMBL:X58776; NID:g29644; PIDN:CAB37831.1; PID:g4467833
R:Harper, J.R.; Prince, J.T.; Healy, P.A.; Stuart, J.K.; Nauman, S.J.; Stallcup, W.B.
J. Neurochem. 56, 797-804, 1991
A:Title: Isolation and sequence of partial cDNA clones of human L1: homology of human
A:Reference number: A50223; MUID:91132183
A:Accession: A50223 :
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1030-1115,'WLC',1118-1176,1181-1257 <RA>
R:Wolff, J.W.; Frank, R.; Muijck, K.; Spiro, R.C.; Relsfeld, R.A.; Rathjen, F.G.
J. Biol. Chem. 263, 11943-11947, 1988
A:Title: A human brain glycoprotein related to the mouse cell adhesion molecule L1.
A:Reference number: A31072; MUID:88298876
A:Accession: A31072 :
A:Molecule type: protein
A:Residues: 'Q',21-36 <WOL>
R:Platzter, M.; Bauer, D.; Drescher, B.
submitted to the EMBL Data Library, March 1995
A:Reference number: B01368
A:Accession: G02506
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-1257 <FLA>
A:Cross-references: FMBL:U52112; NID:g1302657; PIDN:AACS51746.1; PID:g1302658
C:Genetics:
A:Gene: GDB:LLCAM :
A:Cross-references: GDB:120133; OMIM:303350; OMIM:308840
A:Map position: Xq28-qx28
A:Introns: 26/1; 31/1; 66/2; 134/1; 175/1; 232/1; 269/2; 331/1; 375/1; 423/1; 460/2;
/2
C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homolog
C:Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmem
F:1-19/Domain: signal sequence <status predicted <SIG>
F:20-1257/Product: neural cell adhesion molecule L1 <status predicted <MAT>
F:257-314/Domain: immunoglobulin homology <IMV1>
F:532-593/Domain: immunoglobulin homology <IMV2>

Query Match 8.7%; Score 761; DB 1; Length 1257;
Best Local Similarity 24.9%; Pred. No. 2.6e-28;
Matches 283; Conservative 159; Mismatches 422; Indels 274; Gaps 45;

Qy 14 LLSLSPNHLFLAQLIPDPDVERGNDGTPIPTSDNDNSLGTYGSRLRQEDFFPRIVEH 73
| | | | | : | : | : | : | : | : | : |

Db 11 LLUCSP---CLLIQIPEYBHGHWMPVIT-----EQSPRLVFW 48
Qy 74 PSDLVNKGSPATLNKABGRPTPTIEWYKGERVETDKD-----DPASHMLPGS 126
Db 49 PIDDI-----SLKCEASGKPEQVFWTRDGVHFKPKEELGVTVTPSRGSSPTITGN 101
Qy 127 LFLRLIVHGRKSRPDEGVTVVCARNYLGVAVSHNASLEVAILLKDFQNFSD-----VWMA 182
Db 102 SNF-----AQRFGQILYRCFASNKLGTAHSH-----EIRLMABGAPFWKETVKPEVE 149
Qy 183 WGEVAPMEQCPGRPHPTISWKKGDSPLDDKDERITI--RGKILMITTRAKD--AGKYV 240
Db 150 EGEVSVVLPNCPSPAEPLRIYWNOSKILHIKQEDRVYMGKQNLNFAWLTSDHSDYIC 209
Qy 241 -----VGTNMGGERSEVAELTVLERPSFKVPSNLAVTVDDSA-----FKCAR 286
Db 210 HAHPFGTTRIIOKEP---IDLRVKATNSMIDRKPRLLPITNSSSHLVALQOQLVLCEAE 267
Qy 287 GDPVPIVYMKKDGELPKSRYEIRD--DHTLAKIKVTAGDMGSTCVAENMGKAEASATL 345
Db 268 GFTPTIKMLKSGPMPADRVTYQNHKTLQLLVGKEEDGEGYCLAENSGSARATV 327
Qy 346 TVQEPHFHVYKPRQDQVALGRTVYFCBATGNPQPAIFVRRRGSQNLFTSQPSSSRF 405
Db 328 TVEAPVWMLKSGSHLGGESTARLDQVQGRQPEYWRIRG-----IPEVELADQY 382
Qy 406 SVSQGDLITVWQSDGWTYICOTLNAGSIIKAYLEVTDIADRP--PPVIRGPVNT 465
Db 383 RI--RGALLISWQPSDWTMTVQCEARNRGLLANAYIVQV-----PAKILAD--NOT 435
Qy 466 -VAVDG-TFVLSCVATGSPVPTILN-RKDGVLSTQDSRIKLENGVLQIRYALQDGR 522
Db 436 TVAGVSTAYLLKAGAPVPSYQWLEDEGTV--LQDERFFPYANGTGLIGLDQADGSR 494
Qy 523 YTCIASPTSGEATWSAYIEVQRFVYVPPPT-----DNLPLS-----562
Db 495 TFLCAANDQNWITIMANKLVKDATQITQGPASTIEKKGRVITPQASDFSLQPSITWR 554
Qy 563 -----APSKPEYTD-----571
Db 555 GGGRLQELQSDKRYFIEDGRILVHSLDYSQNGSYCASTELDVESRAQLLVSPGP 614
Qy 572 -----VSNNTVLSQWNLNSGATPTYSIIE--AFSMASSGQPTVAENVKYETS 619
Db 615 VPRVLSDHLQTSQVQVWSPADHNAPEKIDIEFEDKEMAPEKMTSLGKVPGQNTS 674
Qy 620 AIRKGLPKNALYFLVRAAYGISDPSQISDPVKTQVLPTSQGVHQRQVRELGNVHL 679
Db 675 TTLKLSPTVTHYFRTVAINKYGQEPGSPSVETVPEAPEKPNFVQVSGNETIMVLI 733
Qy 680 LHNPTVLSSSSIEVHWVDOQSQYIGQIKILYRSGNHGSDWLVETPKPAKNSVPI 739
Db 734 -----TWKPLRN--MKNAPQVQ--FRVQWRQGT-----RGWQEQIVSDP-----FLVVS 776
Qy 740 DLKRGVYWEIKARPPNFQAGSEIKFAKTELEASAPP--QGVTVSKNDGNGTALLVS 797
Db 777 WTSYFVPIEIKVQVNSQKGPQVTVIGYSGEDVPAIPDLGIEIL-----NSSAVLK 832
Qy 798 WQPPEPTDQNGVQEKYV--WCLGNTRYI-----HINK-----TVDSGTSVWIPFLVPIRIS 849
Db 833 WRPDVLAQVHGLRGYNTVYWRGSGQRKSHKHORVWPANTSYVLGSLPTSYSTH 892
Qy 850 VEYASATGAGSKWSE-----PQITLDAGHN-----PVS-----879
Db 893 LEVQAFNGRSGSPASEFTFTPEGVGHPEALHLECGQNTSLLWQPLSHNGVLITGV 952
Qy 880 -----PEDQVSLAQGISDVQKPAFIAGIACAWILWYFSINLYRHRKRNGLTSTYAGI 935
Db 953 LSYPLIDEGGK--QLSFNLKDPGL-----RTNULDLSPGL 987
Qy 936 RKVSPFTPTIPVYQRC--GBA--VSSGGRPGL-----NISPEAPQWLAOTV--PWG 984
Db 988 R-----YRFQLOATTKEGPGAIVREGCTWALSCDNGSATAGENTSVSWPKRG 1041

RESULT 8

150600

neogenin - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: 150600.

R:Vielmetter, J.; Kayem, J.F.; Roman, J.M.; Dreyer, W.J.

J. Cell Biol. 127, 2009-2020, 1994

A:Title: Neogenin, an avian cell surface protein expressed during terminal neuronal d

A:Reference number: A55193; MUID:95105243

A:Accession: 150600

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1443 <VID>

A:Cross-references: EMBL:U07644; NID:641965; PID:AA059662.1; PID:641966

Query Match 8.7%; Score 757.5; DB 2; Length 1443;
Best Local Similarity 23.2%; Pred. No. 4.5e-28;
Matches 378; Conservative 207; Mismatches 681; Indels 361; Gaps 64;

Qy 57 TGSRLQPEPP--RIVEHPSDLVNSKGPATLNKABGRPTPTIEWYKGERVETDKDD 115

Db 9 GSGVVR--TFTPTFLVPEMDLVNSKGVASVIMNCSYCYCTPPKIEWKDCTLLNWSDD- 65

Qy 116 RSHMLPGSLFLPLRVHGRKSRPDEGVTVVCARN--NYLGEAVSHNASLEVAILLKDFRQ 174

Db 66 --RQQLPGSLLNSVWHSKHNKPDSEGYQCVATVSELGSIYSTAKTLVAGL--PFTS 122

Qy 175 NPSDVWVAGSPVMEQCPGRPHPTISWKKGDSPLDDKDERITIRGKILMITTRKSD 234

Db 123 QPZLSVYTKSGSALNCE--VNVDLAPVWRQDQPLSLDDRVFLPSGALLIGNATD 181

Qy 235 AGKYVCGVNMGGERSEVAELTVLERPS-----FKVPSNLAVTVDDSAEFCARGD 288

Db 182 GFTYCRVBSGQPTKYSIEAEALILPDPEEQSLVFTYRQPSLUKTVQARNVYCVAGG 241

Qy 289 PVPVTVMKRDKGEL---PKSRYEIRDDHTLAKIKVTAGDMGSTCVAENMGKAEASATL 345

Db 242 PVPVTVWKLNSHLLTEDSRPALRAGSGSLISDVTEDGVGTICADNENETIEAQL 301

Qy 346 TVQEPHFHVYKPRQDQVALGRTVYFCBATGNPQPAIFVRRRGSQNLFTSQPSSSRF 405

Db 302 AVQVQPEFLKRPANITAEKSDIVFECVTKGPTPTVYKWKNGDVLPSDY-----F 353

Qy 406 SVSQGDLITVWQSDGWTYICOTLNAGSIIKAYLEVTDIADRP--PPVIRGPVN 463

Db 354 RIVKNEQLVGLVKSDEGTQCAENQVGAAGAQLLILDVAIPITLPTSLTATN 413

Qy 464 QTVAGDVTFVLSCVATGSPVPTILNRKQGVLSVTSQDSRIKLENGVLQIRYALQDGR 523

Db 414 DHAL-----PATTPGCUPTAPRDVATLVST-----RPLRL-----443

Qy 524 TCIASPTSGEATWSAYIEVQRFVYVPPPTDNLPLSAPSKPEVIDVSNVTLSWQP 563

Db 444 -----TWK-----TFVSDQ--GDLNYSIFPYTK--GIMRERVNTERP 475

Qy 584 NLMSGATPTSYLIEAFSGAGSNGWQVAVNKVETSA--IKGLKPNALYFLVRAANAYCI 642

Db 480 G-----ETOVMIQNLMPETVTVFRVVAQNKHG 507

Qy 643 SDPSQSDPVKTQDVLPTSQGVHQRQVRELGNVHLNPTVLSSSSIEVHW--TVDOQS 701

Db 508 GESSA--PLX-----VATQPEV--QLGPGAPNRYAGSGP-----SVTVWETPLSGN 552

Qy 702 QTVIGQKYLKPSGANHGESDNLVFSVTRPAKNSVYIPDLRGVNYEIKARPPNFQGA 761

Db 553 GEIQNKLYTKKSGDQ--SEQDQV-----AGLSYITIGLKTYTETSRFVANKHGP 605

Qy 762 DSEIKFAKTELEASAPPQGVTVSKNDGNTALLVSQWPPEDTQNGVQEKYKWLNE 821

Db 988 R-----YRFQLOATTKEGPGAIVREGCTWALSCDNGSATAGENTSVSWPKRG 1041

RESULT 9
S36126
neural cell adhesion molecule L1 - rat
N:Alternate names: nerve growth factor-inducible large external glycoprotein; NILE glycoprotein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S36126; S17655; A60917; A30326
R:Miura, M.; Kobayashi, M.; Asou, H.; Uyemura, K.
FEBS Lett. 289, 91-95, 1991
A:Title: Molecular cloning of cDNA encoding the rat neural cell adhesion molecule L1. The

Qy	1	MKKKVFVWISLLLSNPHLIAQLIPPEVERGCHGRTPIPTSDNDONSJLGYTGR	60
Db	4	MLAVVFLLLCSFCLLIQ-----IPDE-----YKGRH	30
Qy	61	LROQDFPPPIVEH-PSDLIVSKGSPALNCKAGRPPTPIEWKGERVEYDK-----	112
Db	3	VLE---PPVITQSPARKLVFPDDISLCKEAGRPQVEFWRKDGIFHKFKELGVHW	6
Qy	113	DDPSRHMLLPGSGLFLRAIVGKSRPSDEGVYCVARNLGEAVSNHSLVALRQD	172
Db	88	EAPYSGSFTIEGNSYAFQF-----GGITRCASNNIGTAMSH-----EQLVAGA	134
Qy	173	RQNPSSD-----VWVAGGPAWKECPPPHGGPEPTISWKKGSLDDKDERITI-RGKML	227
Db	135	PKMKETKVPVEEGESVLPCCPPTSPAAPIRLVWMSKILIKQDEVSNGGQDLIT	194
Qy	228	TYTKSD-AGITVCGTNGMVER---BSEVALTLEPSPFVRKPSNLAVTDDSAE---	280
Db	195	ANVLSDMSDYIC-NARHPGRTIIQEPIDLRVKPNSMIDKRPFLPTFNSSHLVA	253
Qy	281	-----FKCEARSDPVPVTRNRKDGELPKSRYEIRDH---PLIKRYITAGDMSGYTC	331
Db	254	LQGSGLILCEIAGFGPPTIKWLPSSDPPTDVR-IYQNHNTQLLNGEEDGETCL	312
Qy	332	ARNWYKGAESAATLTVQSPHFVPRQVVALGRTVPOCEATGNPQALPVRREGSN	391
Db	313	ARNLSGASRHHYTVYEAAPWLQPGSHLQPGGTALRQCVQGRPQPEVYWRING-	369
Qy	332	LLFSYQPPSSSFSFSQSDGLIINVRSDGYGIIQZLNVAGSILITKATLETVDIAD	451
Db	370	-MSIEKWNQKTRIED-GLSILSNQPSDWTWQCEARNGHLLANATYIVQL---	423
Qy	452	RPPIYIQGFPWNT-VAYDG-TFLVSCVATGSPVPTILWRKQDGLVTSQDSRIKQENG	509
Db	424	-PARILTD-NUTYNAVSGSTALILCKAGAPSPVSWLDEEGTVLQDERFPFPAHG	480

Qy 510 LQIRYAKLGDGRITCIASPTSGEATWSAYIEVQPGVQPPRP----- 555
Db 481 LGRIQLQANDTGYTFOAANDQNNVITLANLQKREATQITQGPSTIEKRGARVITFOQA 540
Qy 556 -DPNL----- 559
Db 541 SDFPSLQASITWRGDGRGLRGSDKYFIEDGVLKSLDYSDQGVSCVASTLEDVE 600
Qy 560 -----IPSAFSPKPEYD-----VSRNVTVLNQPNLMSGATPTSTIIB-APSHAGSS 606
Db 601 SRAQLLVGSPGPHLESLDRHLKQSQVHLSNSP AEDHNSPIKZDIEFDEKMAPEK 660
Qy 607 WQYVAENKVTSAIKGLKPNAYILFLVRAANAYGLSDPSQISDPYKIQDVLPSQGVQH 666
Db 661 WFLSGKVPQGTSTILKSPVHYTFVTAINKYKGPSPVSTVTPPEAAEKNPVDV 720
Qy 667 KQVQELGNALVHLNPTVLSSSIEVHTVDDQSYQIYKILYRPSGANHGESDNLV 726
Db 721 RGGNGNTNMTV-----TWKPLK-WDWAQIQ-YRQVRLGX-----QETW--- 762
Qy 727 EYVTPARKNSVIPDLRGVWEIKARPFNEPGQADSEIKFATLEBASAPP-QGVTV 784
Db 763 KQVTSDDPLVWNTSTFVPEIKVQAVNCKGQGPQVITGYSGEDPQSPLELDTI 822
Qy 785 SKNDGNGTALVSWQPPEDTQNGMVSQYK-VCLGNTRY-----HINKT-----VDGSTFS 836
Db 823 F-----NSSTVLVWRVDLAQVGLRGVNTYTWKGSQKREKREKSEWVPANTTS 878
Qy 837 WIFPLVPGIRYSVEVAASGAGSGVSE-----PQIQLDAGH----- 875
Db 879 ALLSGLRPTSSYHVEVQAFNGRLGAPASEWTFSTPBGVGHPEALHLBQSDTSLLLHQ 938
Qy 876 -----NPVSPEDVSLAQISDVYKQVPAFVIGAGAACWILLMPSINLY 919
Db 939 PPLSHNGVLTGILLSYHPLDGSKEQLFNLSD----- 971
Qy 920 RHKKKMLGNTSYTAGIRKVPSTF-TPVTIVQGEAV-----SSGGPGLNISEPA 971
Db 972 -PELRNHNVLNPLDQ-----YRFQLATYTGQSGEAVIREGSGTALPQPGFNGSVTA 1026
Qy 972 AQPMLADTW-PWTC 984
Db 1027 GENYSVSVNPREZ 1040

RESULT 10

1669

hor suppressor - African clawed frog

Species: *Xenopus laevis* (African clawed frog)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #start_change 21-Jul-2000

C/Accession: I51669

R: Pierce, C. L., W. E. Reale, M. A.; Candia, A. F.; Wright, C. V.; Cho, K. R.; Fearon, E. R.

Dev. Biol. 165, 654-665, 1994

A>Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the

A/Reference number: I51668; MIMD: 95113183

A/Accession: I51669

A/Status: preliminary; translated from GE/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1427 <P>E

A/Cross-references: EMBL:U10986; NID:606873; PIDN:AAA70168.1; PID:606874

C/Genetics:

A/Gene: XDCCA

Query Match 8.3%; Score 726.5; DB 2; Length 1427;
Best Local Similarity 20.6%; Pred. No. 1.3e-26;
Matches 348; Conservative 201; Mismatches 584; Indels 553; Gaps 55;

Qy 71 VERPSDLIVSKGEPATLNCRAEG-RPTPIEWYKGGREVDKDDPSRHLKPSGLFF 129
Db 43 LSPSDLVTHNGGNNVNLCSAGSDRGAPLKNKDKGVLYNLVIDE-----RQQLPSSGLFI 99
Qy 130 LRIYHKGKSRPOBGVYVCAR-NYLGEAVSHNSLEVA-----ILRDDFRQNFSDVMVWAG 184

Db 100 QNVVSHRRHRPDBGVYQCBASLDSVTIVSRKATKVLVAGLRLI-----SQTESVATVG 154
Qy 185 EPANVQCPGPPHGPPTISWKKDGSPLDKDERIT-----TRGGKIMITTTXSDAG 236
Db 155 DIALACE-ITGEPMTISWQK-----NEEDLVKTPGPRLLVLSGLQLSRQLADGG 208
Qy 237 KYVCVQWNGVGERSEVBLTVLERPS-----FVKRPNSLVVDDSAEFKCEAGDPV 290
Db 209 VYELAKNPGSARVNGEALRLISESGHLRQVFLQPSNVVAILGGQVLAECVSTPT 268
Qy 291 PTWVKWDDGELP-----KSRYEIADDTLIRKVTAGDMSGTCVARNMKAZASLTIVQ 348
Db 269 PLTWQMGDEPPIRTKYSVLGSGNLLISNVTDGAGATVCAATKNENTSFASDLTVM 328
Qy 349 BPPHYVMPKPDQVWALGRVTVQCEATGNPQPAFWRRGSGNLLFSTQPPQSSKFSVS 408
Db 329 VPQQLNHPANLYATESMDIEFBCAVSGKSPSPVKKTKNGEVLPSDY-----FQIV 380
Qy 409 QTGLDILNVOGSDVGIYICQLNVAGSIIITAKYLEVTVIDADRPPVIRQPGVWIVAV 468
Db 381 DGSNLRILGLVSGDGTQYCIANERAGNIQTAYQLIIPD----- 419
Qy 469 DGTFLVSLCATGSPVPTILWRKDGVLSTQDSRIKLENGVLIQIRYAKLGDGTGYTCIAS 528
Db 420 ----- 420
Qy 529 TPGSEANWSATIEVQPGVQPPRPDNLIPSAKSPKPEYDVRNVTVLNQPNLMSG 588
Db 420 -----PAVSSSLPSAPRDVWVPLVSRFVLSNRPPVSK 456
Qy 589 ATPSYIIEAFSHAGSSQWVTAENKETS-----AIGKLPNAYILFLVRAANAY 641
Db 457 GNQITVTV-----FSKQVQREARWNTSQISLIQTVGNLPEETYNFVWAVNGS 509
Qy 642 ISDPSQISDPK-TCQVPLPSQGVHQQVQKRGVNLVHLNPTVLSSSIEVHTVDDQ 699
Db 510 -----PGSSQVGVVQVQLPQVPGVFNQVSVTAPISVLISNDPAPANGP----- 557
Qy 700 QSQYIQYKILYR-SCANHG-ESDWLVETVTPARKNSVIPDLRGVWEIKARPFN 756
Db 558 -----VGTRLCAETFSGRQNIETDGIYR-----LEGAKTETSINVLAKNR 603
Qy 757 EPGQAGSEIKFATLEBASAPPQGVYTKNDGNGTALVSWQPPEDTQNGMVSQYK 816
Db 604 YGPGVSSSEHYVTLSDVPSAMPQNVSLV-ANSGSLVSWLPPPTQNGFITGTRK 661
Qy 817 CLGNETRIHINKTVDGSTFSVIPPLVPGINISVEVAASGAGSGVKS-----EPQIQL 871
Db 662 HXKTTKRGEL-ETLEPNLWLTPLTLENGSQYSQVFAAMTVNGTSPGSDWTATPEMDL 720
Qy 872 D-----AHGNVPSPEDQVSLAQISDVVQVPAFVIGAGAAC----- 907
Db 721 DESQVDPQPSLHVRPLTISIMSWTPPLPNVIVNGIIGTGVSPTATVYDQSKRT 780
Qy 908 -----WILLMPS-----IWLHRRKRN-----GLTSTYA 933
Db 781 YSINLEPPSSSTVILSKAFNAGBGVPLYSATRTSGTVPDMSTPLMPVQGVVALHFD 840
Qy 934 GIR-----KVPSTPTPTVITQRGGEVS-----SGGRGP-LIANI 967
Db 841 AVYVSWADNSYKMKQKTYEVRFTITLRNRTSYASSKYSADTISLSHTVGLKPMTYEF 900
Qy 968 SEPAQAPWLAQTPWNTGNHNDSCCTAGNDSNLTYSY-----PACDIANTNQGLD 1024
Db 901 SWMTYKGRSSSTWNTAH-----ATTETAPTAPDKDITVIRKRPRAVYSGWPIE 954
Qy 1025 -----NQITNMLP-----STVYGVDLSNKNIMKT-----PNSPNLKDGRFVN 1065
Db 955 ANGKIDFLPILDTLKLMLQDDWIMVITIGD-RLTHEILDNLDTATFYRIQARNKGLG 1013
Qy 1066 PSQGTPTPATIQLQISLNNMNGSDSGEKHWLQKQKQGVAPVQINIVQNKLNKD 1125
Db ----- 1125

124 EPMHVREGSDILACNRPPVGLPPPIIPWMDNAQRPLQDSERVS--QLGQNLIFSNVQPED 182

QY 235 AGI--YVCVG--TNMVGSESVAEVLTLERSPVKRPSNLAIVVDGSA-----F 281

DB 183 TREDYICARNEHIIQKQKQPSIKSVKSTPKPTERRPITLPLMGSTSNKVELRGNLL 242

QY 282 KCBARQDPVPTVWRKNDGLKSPKSEYERD-DHTAIKRVTAGMGSTYCVAAENSGAE 340

DB 243 ECTAAGLPTPTVIRKIBGSELLPNTKPFENKKTILAEIVSDSGKYKCTAINTLGTH 302

QY 341 ASALTVPQEPHPTVVKPRQDVVALGRVTPQCEATGNQPAIFVRRSGSNLFSYQPPQ 400

DB 303 HVISVTKRAFPITAPNLLVLSPOEDGLICRANRKPFSISLWNGVPIAT-----APE 35

QY 401 SSSRFVSYSQDGLTIL-TNVRGSDWGYICQLTNVAGSIITKALEVTDVIADRPVPIVQ 459

DB 359 DFRS-----KYDGTITFSAVERSSAVTQCMASNEYTLIANLAFV-----NYLAE-PPRIL- 410

QY 460 GPWQV--TVAVDQFVLSCVARGSPPTILMRKQGLVLS-TQDSRIQKLEWGLVIRYAK 516

DB 411 -PANKLVQIADSPALDCATFSGKPELEWFK-GYKSLRGNEVTFHDGDLIEPIAQ 468

QY 517 LGDTGTCTIATSPGSEATWSAIVLEQEGFVQPPR-----PTDPNLI 560

DB 469 KDSGTGTYTCARNLIGKTQNEVLQKVPDPMIIPKQPKVIRQSQAQSFCEVICKDPTLI 528

QY 561 P-----SA 563

DB 529 PTIVLKNDDNPLPDERFLVGKDNLIIMNVTDKDDGTTCYIVNVTLDVSASAVLTVA 589

QY 564 PSKP-----EYTVSRNVTLSWQPNLNSGATPTSYIIAPFS--HAGSSWQ 608

DB 589 PPTAFITARPNPDLIELTGQLERSLELFWGEGNNSPITNFVIEYEDGLHPG-VWH 647

QY 609 TVAENKTESAISGLKPNATILFLVRANAYGSDSPISDPVKTQDVLPTSQGDYHK 668

DB 648 QTQVEPGSGTIVLKLSPTVIVSPFLVAINVEIKSGSPSEBQYLTKSANDPNPFWQ 707

QY 669 VQRELGAVNLBNPTVLTSSSIVYHWVDQSQGTQGLIKLYRPSGANGSGDLVFEY 728

DB 708 IGSEPNLNVITWESLKGFGSGNGLQ-----TKVSWQKDV--DDEN----- 747

QY 729 RIFAKNSVIPDLK-----GWNYLEKARFPNFPNGQADSEIKFAKLEBASAPPO 780

DB 748 ----TSVANNKSVISGTPPTVPEIKVQALNOLGATAPPSEYVGHSGEDLPWAPG 802

QY 781 GVTYSKNDGNTALLWSWPPPEDTQNGWQYKVV-----CLGNETRIHMK--TVDG 832

DB 803 NVQV--HYVNTLARKHMDPVLKSRVGHGLQYKVVYKQSLGSRKKRVEKKLILFRG 860

QY 833 STFSVIPFLVPGIRVSEVAATGAG-----SGVKSBPQIV-----LDA- 873

DB 861 KRTGLFGLPELTYSSTKLNRVNVKNGSGPASPDKFKTPGCVSPSPFKLTNPIDSL 920

QY 874 ----HGNVSPSE-----DQVSLAQIISDV-----VKQPAFIAGAACWILMYSIWL 918

DB 921 TLWGSPTPRNGVLTILKPLNTINHELGAPEVIRIAPNESS-----LILNMLN 971

QY 919 YRRHK--RNLSTYAGIRKVPSTFTPTVYQRG--GEAVSSGG--RPLGLNISIPA 971

DB 972 YSTRYKFIAGTSVSGSGQITEE--AVTIDEGALRPVAGKGVQVLPRIPIRVNTIA 1028

QY 972 AQPNLADTWPTGNNHDCISCTAG-----NGNSDNLITYSRP-----1012

DB 1029 AETVAINSEYEPDPAINTVETVYAGSKEDWKLVNGSRFVLKGLTPGATKVVKG 1088

QY 1013 ADCIANYNQNLNQTNLMLPESTVYGDVLSKNINEMKTNNSNLKGRGFVNFSGQPT 1072

DB 1089 ABELGFSFSSDLEFETGPAMAR-----QVQDITQ-----GWF--GLMCA 1127

QY 1073 YATPOLIGQNSNMNNGSGDSEKXHWPLGQOQGEVAPQYVNIQENLKNKYDRANDTV 1132

DB 1128 VALLILILILVIFCIRRNKGG-----KTPYKEX--EDAHADPEI 1163

Qy	68	PRIVEHPESDLIV-----SKGEPATKACAGKGPPTFENKYGKEGVEVDKDDPSRHEMLPS	124
Db	26	PPVFPKPSNPISFVPGSGDKITLLCEARGNPSPHYKWLQSGSDITSDLS----HYTKLWG	81
Qy	125	GLSLFLVLRHVKRSPDGBGVTVCAVRYLGEAVSNASLEVALRDOFPNQFSDVMVWAG	184
Db	82	GNL-----TVINPKNWIDGSTGTCFATNSLGTISREALKLPATLENFNRMSRVSVRBG	137
Qy	185	EPAMVDCOPPPHGPPEPTISWKKDGPL-----DDKDKIRIAGGKLMTITRSDSGKVCV	241
Db	138	QGVWLCCPPPHSGELSYAWFENKPSFVEDRNRFSVGTGLGTLAKVEYSDGVNCTW	197
Qy	242	GTINW-----GERSEVAEVZVLEKPSFVKRPSNLAVTVDGSAEF	281

Db 198 VYSTVITNARVLGSPPTLVLRSDDGVMGEYEPKIE-----LQPPETLPAARGSTVWL 247
 Qy 282 KCBAEGDPVTVYMKRKGDELPSKSYEIRD-DHTLIRKVTAGDMGTCVCAENVMGAE 340
 Db 248 ECFALGNPVPQINWRSGDGMPPPTIKILKRFNGLVNPQEDQSTGYCIAENSGKVN 307
 Qy 341 ASATLTIVQEPHFVYKPDQVVALGRVTTPCCATGNQPAIFVRRSGSNLFTSQPPQ 400
 Db 308 ARGRLITIAKTVQLLQDLVETAVEDSLVWECRASGKPKPTWHLKNGDALV----- 360
 Qy 401 SSSRFVSQSGDLITINQSRSDGVYICQTLNAGSITIKAYLEVVIDIADPPVIRQ 460
 Db 361 -EERIQI-ENGALITIANLVSDGMPQCIANHGSLYSASLKA-----VLASAPD--FSRN 413
 Qy 461 PVNQ--TVAGDVTFLVSCATGSPVPTILWRKDGVLSTQDSRIQLENGVLQIRYAKIG 518
 Db 414 PMKIMQVQWGLSVILDCPKSPASPRALSFVKMGDVTVRQD-ARISLNDGGLKIMVTKA 472
 Qy 519 DTGRTICTASIPSGEATWSAIEVQEPGVPPPPRDTNPLIPS----- 562
 Db 473 DAGITICTAENQFGKANGTTLVYTE-----PRTILAFSNDMVAGSEIILPCQV 523
 Qy 563 ----- 562
 Db 524 QHDLPLIDFAMVYFNGLTDFKKGDSHEFKVGGSSGDLMINIQLKHSKTVCMVQTGV 583
 Qy 563 -----APSKPE--VTDVSRNTVLWQPNLNSGATPTIIEAFSHAGSS 606
 Db 584 DVSAAELVLVSGSGPPENVKVEIDITDTAQLSWTEGDSHSFVISTAYQAQPTPS-VG 642
 Qy 607 WQ--TVAENV--KTETSAIKGLKPNAILFLVRAANAYGISDQSQISDPVKTDVLP 661
 Db 643 WQVQVTPVLDGKTHATVVELNFWVEYEFVAVSNKGGGSPLESEKVRTEAAPE- 701
 Qy 662 QGVDRQVQRELGNALVHLNPTVLSSSIEVHTVDOOSQIY----GTYLIRPSPA 716
 Db 702 -VAPSEVSGGG-----SRSELVTN--DPPELQNGGFGTVFVAPRGLV 745
 Qy 717 NHGESDMLVFEVPT-----AKNSVIPDLRGVNYEIKARPPFNEFGADSEIKAFKT 770
 Db 746 ----TWIQTIVTSPDNPRVFNESIVP--FSPTEYKGVYNNNGEGPSPVTVFS 797
 Qy 771 LEEAPSPAGQVTVSKNDGNGTALVSWQPPEDTQNGMVQYKVV--WCLGNETRYHNK 828
 Db 798 ABEETVAPS--HISAHLSSEIEVSWNTIPWKSNGRLGLGYEVRTWNGGEESSSKV 855
 Qy 829 TVDGSFTSVIPLFVPGIRSYEVAASGAGSG 861
 Db 856 KVAGNQTSAVLRLGLKSLNAYTAVRAYNTAGAG 888

RESULT 14

A53449

plasmacytoma-associated neuronal glycoprotein PANG - mouse

C:Species: Mus musculus (house mouse)

C:Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 24-Sep-1999

C:Accession: A53449

R:Connelly, M.A.; Grady, R.C.; Mushinski, J.F.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 91, 1337-1341, 1994

A:Title: PANG, a gene encoding a neuronal glycoprotein, is ectopically activated by int2

A:Reference number: A53449; MUID:94151325

A:Accession: A53449

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1028 <CON>

A:Cross-references: GB:L01991; NID:g200056; PIDN:AAAL7403.1; PID:g200057

C:Superfamily: catonin; fibronectin type III repeat homology; immunoglobulin homology

C:Keywords: glycoprotein

Query Match 8.0%; Score 695.5; DB 2; Length 1028;

Best Local Similarity 25.5%; Pred. No. 2.4e-25;

Matches 238; Conservative 125; Mismatches 361; Indels 209; Gaps 28;

Qy 58 PRIVEPSSOLIV-----SKGEPATLNCAGRPPTIENYKGEVETDKDPSHRMLPS 124
 Db 26 PVTIKPSNSIIPVDSKDKITLNCAGRNPSPHYKWLNGSDIDSLD-----HYIKNG 81
 Qy 125 GSLFFLRVHVRKSRPDGQYVYVARNVLEGVSHASLEVALRDDFPQNSDVMVWAG 184
 Db 82 GNL-----IVIPNNWQSGYQCFATNSLGTIVSREAKLQPAYLENTKRMESTVSRE 137
 Qy 185 EPANWECOPPPHPEPTIISWKGDSPL-----DDEKERTIRGKMLITTYKSDAGYV 241
 Db 138 QGVLLCGPPPHSGELSYANVNEPSPVEDSRRLVSGQTHLITAKVPSDQVNTVCV 197
 Qy 242 GTNV-----GERESYAEVLITLERPSFKVRPSNLAIVTDSAEF 281
 Db 198 VYSTVITNARVLGSPPTLVLRSDDGVMGEYEPKIE-----VQPPETLPAARGSTVRL 247
 Qy 282 KCBAEGDPVTVYMKRKGDELPSKSYEIRD-DHTLIRKVTAGDMGTCVCAENVMGAE 340
 Db 248 ECFALGNPVPQINWRSGDGMPPPTIKILKRFNGLVNPQEDQSTGYCIAENSGKVN 307
 Qy 341 ASATLTIVQEPHFVYKPDQVVALGRVTTPCCATGNQPAIFVRRSGSNLFTSQPPQ 400
 Db 308 ARGRLITIAKTVQLLQDLVETAVEDSLVWECRASGKPKPTWHLKNGDALV----- 360
 Qy 401 SSSRFVSQSGDLITINQSRSDGVYICQTLNAGSITIKAYLEVVIDIADPPVIRQ 460
 Db 361 -EERIQI-ENGALITIANLVSDGMPQCIANHGSLYSASLKA-----VLASAPD--FSRN 413
 Qy 461 PVNQ--TVAGDVTFLVSCATGSPVPTILWRKDGVLSTQDSRIQLENGVLQIRYAKIG 518
 Db 414 PMKIMQVQWGLSVILDCPKSPASPRALSFVKMGDVTVRQD-ARVSLNDGGLKIMVTKA 472
 Qy 519 DTGRTICTASIPSGEATWSAIEVQEPGVPPPPRDTNPLIPS----- 562
 Db 473 DAGITICTAENQFGKANGTTLVYTE-----PRTILAFSNDMVAGSEIILPCQV 523
 Qy 563 ----- 562
 Db 524 QHDLPLIDFAMVYFNGALTDFKKGDSHEFKVGGSSGDLMINIQLKHSKTVCMVQTGV 583
 Qy 563 -----APSKPE--VTDVSRNTVLWQPNLNSGATPTIIEAFSHAGSS 606
 Db 584 DVSAAELVLVSGSGPPENVKVEIDITDTAQLSWTEGDSHSFVISTAYQAQPTPS-VG 642
 Qy 607 WQ--TVAENV--KTETSAIKGLKPNAILFLVRAANAYGISDQSQISDPVKTDVLP 661
 Db 643 WQVQVTPVLDGKTHATVVELNFWVEYEFVAVSNKGGGSPLESEKVRTEAAPE- 701
 Qy 662 QGVDRQVQRELGNALVHLNPTVLSSSIEVHTVDOOSQIY----GTYLIRPSPA 716
 Db 702 -VAPSEVSGGG-----SRSELVTN--DPPELQNGGFGTVFVAPRGLV 745
 Qy 717 NHGESDMLVFEVPT-----AKNSVIPDLRGVNYEIKARPPFNEFGADSEIKAFKT 770
 Db 746 ----TWIQTIVTSPDNPRVFNESIVP--FSPTEYKGVYNNNGEGPSPVTVFS 797
 Qy 771 LEEAPSPAGQVTVSKNDGNGTALVSWQPPEDTQNGMVQYKVV--WCLGNETRYHNK 828
 Db 798 ABEETVAPS--HISAHLSSEIEVSWNTIPWKSNGRLGLGYEVRTWNGGEESSSKV 855
 Qy 829 TVDGSFTSVIPLFVPGIRSYEVAASGAGSG 861
 Db 856 KVAGNQTSAVLRLGLKSLNAYTAVRAYNTAGAG 888

RESULT 15

A39640

neural cell adhesion molecule Nr-CAM precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A39640; S16451

00 77Z 181770Z 170617FIBIRYQNRDLSIRAFSEIAGHSCEDDFATAPG 020

Search completed: January 22, 2001, 12:27:16
Job time: 2113 sec

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:28:41 ; Search time 162.41 Seconds
(without alignments)
328.290 Million cell updates/sec

Title: US-09-540-245a-18
Perfect score: 8724
Sequence: 1 MKWKVPLVWISLLSLSPN.....VLGGTERGEDNNELEETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Search: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	DB ID	Description
1	764.5	8.8	1260	1 CAML_MOUSE	P11627 mus musculus
2	761	8.7	1257	1 CAML_HUMAN	P32004 homo sapien
3	760	8.7	1493	1 NEOL_MOUSE	P97798 mus musculus
4	757.5	8.7	1443	1 NEOL_CHICK	Q90610 gallus gall
5	755	8.7	2012	1 DSCA_HUMAN	O60469 homo sapien
6	746.5	8.6	1377	1 NEOL_RAT	P97603 rattus norv
7	739.5	8.5	1259	1 CAML_RAT	Q05695 rattus norv
8	725.5	8.3	1461	1 NEOL_MOUSE	Q92859 homo sapien
9	701.5	8.0	1284	1 NRCA_CHICK	P35331 gallus gall
10	679.5	7.8	1447	1 DCC_MOUSE	P70211 mus musculus
11	667.5	7.7	1447	1 DCC_HUMAN	P43146 homo sapien
12	653.5	7.5	1040	1 AXOL_HUMAN	Q02246 homo sapien
13	652	7.5	1040	1 AXOL_RAT	P22063 rattus norv
14	633.5	7.3	1036	1 AXOL_CHICK	P28695 gallus gall
15	627	7.2	1239	1 NRG_DROME	P20241 drosophila
16	596.5	6.8	1018	1 CONT_HUMAN	Q12860 homo sapien
17	589.5	6.8	1010	1 CONT_CHICK	P14781 gallus gall
18	584	6.7	1912	1 PTFP_HUMAN	P23468 homo sapien
19	583	6.7	2029	1 LAR_DROME	P16621 drosophila
20	573	6.6	1020	1 CONT_MOUSE	P12960 mus musculus
21	549	6.3	1266	1 NGCA_CHICK	Q03696 gallus gall
22	549	6.3	1897	1 PTFP_HUMAN	P10596 homo sapien
23	505	5.8	1070	1 PTRK_HUMAN	P13308 homo sapien
24	500.5	5.7	3707	1 PGBM_MOUSE	Q05793 mus musculus
25	500	5.7	1091	1 NCAL_MOUSE	P13590 gallus gall
26	497.5	5.7	1115	1 NCAL_HUMAN	P13595 mus musculus
27	495	5.7	4393	1 PGBM_HUMAN	P98160 homo sapien
28	490	5.6	1051	1 PTRK_CHICK	Q91048 gallus gall
29	483.5	5.5	853	1 NCAL_BOVIN	P18186 bos taurus
30	478.5	5.5	848	1 NCAL_HUMAN	P13591 homo sapien
31	466	5.3	761	1 NC2A_HUMAN	P13592 homo sapien
32	465	5.3	858	1 NC2A_RAT	P13596 rattus norv
33	460	5.3	1092	1 NC2A_XENLA	P36335 xenopus lae

34	450	5.2	837	1 NC2A_MOUSE	O35136 mus musculus
35	449.5	5.2	1088	1 NCAL_XENLA	P16170 xenopus lae
36	448.5	5.1	725	1 NC2A_MOUSE	P13594 mus musculus
37	426.5	4.9	837	1 NC2A_HUMAN	O15394 homo sapien
38	422.5	4.8	811	1 FS22_DROME	P34083 drosophila
39	422.5	4.8	873	1 FS21_DROME	P34082 drosophila
40	415.5	4.8	1906	1 KMLS_CHICK	P11799 gallus gall
41	401.5	4.6	2481	1 UNS2_CAEEL	Q06561 caenorhabdi
42	398.5	4.6	1913	1 KMLS_HUMAN	O15746 homo sapien
43	390.5	4.5	1274	1 MYPC_HUMAN	O14896 homo sapien
44	387.5	4.4	1270	1 MYPC_MOUSE	O70468 mus musculus
45	379	4.3	1271	1 MYPC_CHICK	O90688 gallus gall

ALIGNMENTS

RESULT 1
CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=BRAIN.
RX MEDLINE=88318924; PubMed=3412448;
RA Moos H., Tacke R., Scherer H., Teplow D., Frueh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
superfamily with binding domains similar to fibronectin.";
RL Nature 334:701-703(1988).
CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
TO AKNOW ON NEURONS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; X12875; CA931368.1; -.
DR PIR; S05479; S05479.
DR HSP; P20241; ICFB.
DR MGD; MG1:96721; L1CAM.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; Ig; 6.
DR PRINTS; PR00014; FNTYPEII.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Signal; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1124 1146 POTENTIAL.
FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.

```

RESULT 2
CAML_HUMAN
ID CAML_HUMAN. STANDARD; PRT; 1257 AA.
AC P32004;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1 OR MIC5.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN
RP [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=92031698; PubMed=1932117;
RA Kobayashi M., Miura M., Asou H., Uyemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of

```

RT different origin.";
 RL Biochim. Biophys. Acta 1090:238-240(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rosenthal A., Coutelle O., Drescher B.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.M.
 RX MEDLINE=92329299; PubMed=1627459;
 RA Reid R.A., Hemperly J.J.;
 RT "Variants of human LI cell adhesion molecule arise through alternate
 RT splicing of RNA.";
 RL J. Mol. Neurosci. 3:127-135(1992).
 RN [4]
 RP SEQUENCE OF 353-1176 FROM N.A.
 RX MEDLINE=92020233; PubMed=1923824;
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
 RT "PCR walking from microdissection clone M54 identifies three exons
 RT from the human gene for the neural cell adhesion molecule LI
 RT (CAM-LI).";
 RL Nucleic Acids Res. 19:5395-5401(1991).
 RN [5]
 RP SEQUENCE OF 332-371 FROM N.A.
 RX MEDLINE=90353957; PubMed=2387585;
 RA Djabali M., Mattai M.-G., Nguyen C., Roux D., Demengeot J.,
 RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
 RT "The gene encoding LI, a neural adhesion molecule of the
 RT immunoglobulin family, is located on the X chromosome in mouse and
 RT man.";
 RL Genomics 7:587-593(1990).
 RN [6]
 RP SEQUENCE OF 1030-1257 FROM N.A.
 RX MEDLINE=91132183; PubMed=1993895;
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
 RA Stallcup W.B.;
 RT "Isolation and sequence of partial cDNA clones of human LI: homology
 RT of human and rodent LI in the cytoplasmic region.";
 RL J. Neurochem. 56:797-804(1991).
 RN [7]
 RP SEQUENCE OF 20-36.
 RX MEDLINE=88298876; PubMed=3136168;
 RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
 RA Rathjen P.G.;
 RT "A human brain glycoprotein related to the mouse cell adhesion
 RT molecule LI.";
 RL J. Biol. Chem. 263:11943-11947(1988).
 RN [8]
 RP VARIANT HSAS TYR-264.
 RX MEDLINE=94004956; PubMed=8401576;
 RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnai D.;
 RT "A missense mutation confirms the LI defect in X-linked hydrocephalus
 RT (HSAS).";
 RL Nat. Genet. 4:331-331(1993).
 RN [9]
 RP VARIANT HSAS/MASA LEU-1194.
 RX MEDLINE=95187172; PubMed=7881431;
 RA Fransen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
 RA Willems P.J.;
 RT "X-linked hydrocephalus and MASA syndrome present in one family are
 RT due to a single missense mutation in exon 28 of the LICAM gene.";
 RL Hum. Mol. Genet. 3:2255-2256(1994).
 RN [10]
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
 RX MEDLINE=95004608; PubMed=7920659;
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
 RA Paterson J., Metzberg A., Ionescu V., Temple K., Kenwick S.;
 RT "X-linked spastic paraplegia (SPGL), MASA syndrome and X-linked
 RT hydrocephalus result from mutations in the LI gene.";
 RL Nat. Genet. 7:402-407(1994).
 RN [11]
 RP VARIANTS MASA GLN-210 AND ASN-598.
 RX MEDLINE=95004609; PubMed=7920660;
 RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
 RA Reyniers E., Korn B., Poustka A., Wilson G., Schrander-Stumpel C.,
 RA Winter R.M., Schwartz C., Willems P.J.;
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene
 RT LICAM.";
 RL Nat. Genet. 7:408-413(1994).
 RN [12]
 RP VARIANTS HSAS/MASA S-9; S-121; X-309; F-768; L-941 AND C-1070.
 RX MEDLINE=95282776; PubMed=7762552;
 RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
 RA Holmberg S., Wadelius C., Kenwick S.;
 RT "New domains of neural cell-adhesion molecule LI implicated in
 RT X-linked hydrocephalus and MASA syndrome.";
 RL Am. J. Hum. Genet. 56:1304-1314(1995).
 RN [13]
 RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
 RX MEDLINE=96153146; PubMed=8556302;
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
 RT retardation, adactyl thumbs, spastic paraparesis and hydrocephalus
 RT due to mutations in one single gene, LI.";
 RL Eur. J. Hum. Genet. 3:273-284(1995).
 RN [14]
 RP ERRATUM.
 RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
 RL Eur. J. Hum. Genet. 4:126-126(1996).
 RN [15]
 RP VARIANTS HSAS/MASA/SPGL SER-179 AND ARG-370.
 RX MEDLINE=96057511; PubMed=7562969;
 RA Ruiz J.C., Cuppens H., Legius E., Frys J.-P., Glover T., Marynen P.,
 RA Cassiman J.-J.;
 RT "Mutations in LI-CAM in two families with X linked complicated
 RT spastic paraplegia, MASA syndrome, and HSAS.";
 RL J. Med. Genet. 32:549-552(1995).
 RN [16]
 RP VARIANTS HSAS CYS-194 AND LEU-240.
 RX MEDLINE=97083370; PubMed=8929944;
 RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
 RA Engel W., Schwilger E., Gal A.;
 RT "Five novel mutations in the LICAM gene in families with X linked
 RT hydrocephalus.";
 RL J. Med. Genet. 33:103-106(1996).
 RN [17]
 RP VARIANTS HSAS Q-184; Y-439-T-443 DEL; C-784 AND L-936-L-948 DEL.
 RX MEDLINE=97338664; PubMed=9195224;
 RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
 RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
 RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
 RT "Nine novel LI CAM mutations in families with X-linked
 RT hydrocephalus.";
 RL Hum. Mutat. 9:512-518(1997).
 RN [18]
 RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
 RX MEDLINE=98180721; PubMed=9521424;
 RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
 RT "Multiple exon screening using restriction endonuclease
 RT fingerprinting (REF) detection of six novel mutations in the LI cell
 RT adhesion molecule (LICAM) gene.";
 RL Hum. Mutat. 11:222-230(1998).
 RN [19]
 RP VARIANT CRASH PRO-632.
 RX MEDLINE=98112489; PubMed=9452110;
 RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
 RA Willems P.J.;
 RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
 RL Hum. Mutat. Suppl. 1:S284-S287(1998).
 RN [20]
 RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
 RX MEDLINE=98415726; PubMed=9744477;
 RA Daugler-Weber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
 RA Saugier A., Henocq A., Heron D., Jouveaux P., Odent S., Manouvrier S.,
 RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
 RT "Identification of novel LICAM mutations using fluorescence-assisted
 RT mismatch analysis.";

AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.

-1- TISSUE SPECIFICITY: WIDELY EXPRESSED.

-1- DEVELOPMENTAL STAGE: EXPRESSED UBICUOUSLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E16.5.

-1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

-1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

-1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO TUMOR SUPPRESSOR PROTEIN DCC.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Y09355; CAA07027.1; .

DR HSSP: P02751; 1TTG.

DR MGD: MGI:1097159; NBO1.

DR INTERPRO: IPRO01777; .

DR INTERPRO: IPRO03006; .

DR PFAM: PF00041; fn3; 6.

DR PFAM: PF00047; ig; 4.

DR PRINTS: PR00014; FWTPEI1.

KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;

KW Alternative splicing.

FT SIGNAL	1	36	POTENTIAL.
FT CHAIN	37	1493	NEOGENIN.
FT DOMAIN	37	1136	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	1137	1157	POTENTIAL.
FT DOMAIN	1158	1493	CYTOPLASMIC (POTENTIAL).
FT DOMAIN	78	147	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	177	239	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	274	338	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	366	428	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	467	564	FIBRONECTIN TYPE-III.
FT DOMAIN	567	660	FIBRONECTIN TYPE-III.
FT DOMAIN	661	760	FIBRONECTIN TYPE-III.
FT DOMAIN	766	860	FIBRONECTIN TYPE-III.
FT DOMAIN	881	981	FIBRONECTIN TYPE-III.
FT DOMAIN	982	1083	FIBRONECTIN TYPE-III.
FT DOMAIN	1149	1153	POLY-VAL.
FT CARBOHYD	84	84	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	221	221	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	337	337	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	501	501	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	520	520	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	670	670	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	746	746	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	940	940	N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC	442	461	MISSING (IN ISOFORM 2).
FT VARSPLIC	863	878	MISSING (IN ISOFORM 3).
FT VARSPLIC	1086	1096	MISSING (IN ISOFORM 4).
FT VARSPLIC	1279	1331	MISSING (IN ISOFORM 5).

SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0B CRC64;

Query Match 8.78; Score 760; DB 1; Length 1493;

Best Local Similarity 23.24; Pred. No. 1.1e-26;

Matches 385; Sensitive 220; Mismatches 608; Indels 446; Gaps 72;

QY 70 IVEPHSLIVSNGEPATLNCKAERPTPIENYKGERVETDKDPKSRHLLPSGSLF 129

Db 67 LVE-PVDTLSVSGSSVILNCATSEPSNPIENKDKDTFLNLESD--RQLQLDPSLFI 122

QY 130 LRIHVKKSRPDEGVYCVAR-NYLGEASVSNASLEVALRDFRQNPDSVMVWGEPAV 188

Db 123 SNIVESKBNKPDGGTQCVATVNLGTIVSRITAKIVAGL-PRFTSQPEPSVTVGNAL 181

QY 189 MECQPGPHPEPTISWKKGSPLDKDERITIRGKLLMTTYRKSAGKYTCVGTINWGE 248

Db 182 LNCE-VNADLVFFVWQEQQLLEDRIYVLPGLSTVISMATGDSGLGYRIVSGPP 240

QY 249 RESEVALILHERPS-----FYKPSNLAIVTVDGSAKFCARGDPVTIVWRKDGEL 302

Db 241 KFSDEAEILVQDPEEIVDLVFLMPSGSMKVTQGSAPLPCVSGSLPAPVWNNKEVL 300

QY 303 ----PKSRVIEKDDHTLKIRKVTAGDSMTTCVAENMWKASATILVQEPFFHVPRK 359

Db 301 DTSSSGRVLVLAGGLCEISDVTEDAGTFFCIADNGKNTVEAQLTQVPPGFLQAPN 360

QY 360 QVVALGRTVTPQCATENPQAPFVWRBEGSNLLFSYPPQSSSRFSVSQGLDITINQV 419

Db 361 IYAESMDIVFCEVTKPPTPKVWKNGDVTI-----PSDNFKVKEHNLVLGLV 412

QY 420 RSDVQYITQCLNVAGSITKALVET--DVIAIDPPFVIGQFVGNVAVDGTIVSLCY 477

Db 413 KSDGEPTQCIADNGWQAGAAQLILHEDVAITPLP-----TSLTATTDILAP 463

QY 478 ATGSPVPTILARKGQVLSQTSRIKQLNGVLIQVIRLADGTGTICATSPGSE-ATW 536

Db 464 ATGTPGLSAPPDVAASLVST-----RPIKL--TWRTPSDEGDNLT 504

QY 537 SATIEVQFQFVQVPPRPPTDPLNIPSAPEVDTSRNVTLSQNLQSNAGTPTIYII 596

Db 505 SVPTIKK--GVDRNVENT-----SQGEMQVT----- 530

QY 597 EAPSHAGSSGQVTAENKTESAKGLKFNALYLFVRAANAYGIDSPQISDVPKVTQ 656

Db 531 -----IQNMNPATVYIKVMAQNGK-SGESSAPLREVTPQ 565

QY 657 VLPTSGQVDEHQVQRELGNVLEHNLPLVSSLSIEVH-VTDQSQYIGVLYILPPSG 715

Db 566 EV-----QLGPPAPIRATASPT-----SITVWETPLSGNGEIQNLYIKMG 611

QY 716 ANHGESDMLVFEVETPAKNSVYIPDLAKGVWIEIKARPFPEQAGDSIEKATLEAP 775

Db 612 TDK-EQDIDV-----SSSETINGLKITYTSFRVVAITNKHGFGVSTQVAIVRLSDVP 664

QY 776 SAPPGQVTSKNDGNTALLVSWQPPEDTQNGWQVETKVCNLGNETRIHINKT-VDGS 834

Db 665 SAAPQKSLGVH--NKSATVIEHQPPSGTINQGIQTKYTRKASKSDVETVLGTQ 722

QY 835 FSVVITVLPISIRYSEVAASGAGGSVSE-----PQFIQDAGNCPV 878

Db 723 LSQLIGELGRDTEINFRVALVYNGTSPATDLAETPSDLDRTVPE-VPSSSLVRL 781

QY 879 -----SPEDQVSLAQISDVWQPAFIAGAACWILMFSWILYRHRKKNGLT 929

Db 782 VTSIVSWVTPENO-----NIVWGLATGYIGSPHAQTIKD--YKQR----- 823

QY 930 SITAGIKRV-PSPFTPIYTV-TQRGEAV-----SGSGRP-----GLNISEP-AAQPV 975

Db 824 -TYTINLQPSSTVITLKAFNNVGGIPLYESAVTRPHITDSEVDFVNAIPVVPD 881

QY 976 LADTWPNIG-----NNHDCSISCTAGNCSNLSLITSPACDIANTWQOLDKQTNL 1030

Db 882 PTPMPPVGVQASILSHDITRTW-----ADNSLPKHKQID--SRITV--EWNK- 929

QY 1031 MLPESTVYGVDLSKNINMKMTFNSP-----LKDGRFVNPSPQPTPIATYQLQSN 1082

Db 930 -IPANTYKXNAN-ATLSLYVLGLKPNLTLYEFSVWTKGRASSTMTAGATFELVPTS 987

QY 1083 LSNKNNGSGDSGEX-----HWKPLQOQKQV-----APQVNIYED--NKLN 1123

Db 988 PPKDVTVSKSGKRTIIVWQPPSEANGIKGIIYITSDYNAEIDHWIEFWVGRLE 1047

QY 1124 KQYRANDVTPPTPIYQNSVDQNTGG-----SYNSDR--GSSTSGSGHKKG 1168

```

RESULT 4
NEOLCHICK
ID NEOLCHICK STANDARD; PRT; 1443 AA.
AC Q90610;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Arcohsauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
MEDLINE-95105243; PubMed-7806578;
Violeten J.G., Roman J.M., Dreyer W.J.;
"Neogenin, an avian cell surface protein expressed during terminal
neuronal differentiation, is closely related to the human tumor
suppressor molecule deleted in colorectal cancer.";
J. Cell Biol. 127:2009-2020(1994).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
CC AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-II-TYPE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstati
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no co
CC modified and this statement is not removed. Usage by and for comm

```

```

Query Match      8.74; Score 757.5; D8 1; Length 1443;
Best Local Similarity 23.24; Pred. No. 1.3e-26;
Matches 378; Conservative 207; Mismatches 681; Indels 361; Gaps 64;

Qy   57 TGSRLRQEDEPPP-RIVEPHSLVSKSGPATLKAGRPPTPIEWYTAGSERVETKDDP 115
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    9 TGSIVR--TFPTFTVLVEEDMLSVRGASYVMNCSSCTPEPPKIKSWKGQTLNLNLSVD 65

Qy   116 RSHRMLLPGSGLFFLIRVHGKSRPDGGVVVCAR-NYLGVASRNASLEVALLRDGRFQ 174
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    66 --RRQLLPDGSLLINSVSHKKNRPDGEGYQCVAIVESLGISVRYAKILVAGL-PRFTS 122

Qy   175 NPDSVNVAVGAPAMWECQPPHGPPEPTIISWKKGSPLDOKDERITIRGKLMIITRKSD 234
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    19 QELSESVTGNSAINLACE-VMDLIAPFRWEQDRPLSLDLDFWLKPSGALLIGNATDIO 181

Qy   235 AGRTVCVGTVNWGESERAEALTVLERPS-----FYWRKSNLAIVTDDSAEKKCARGD 288
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    182 GGFTACVIESGGTPKYSEALKILPOPEEPQSILVFYRQPSLLITVQTGNANVPFCVAGGF 241

Qy   289 PVPYVNRKDDGEL---PKSYRIEDDHLLKRVKTAGDMSGYSTCAVNWNWGAKASALT 345
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    242 PTYPYVYNNGEELLTEDESFERALARAGSGLLISUVEDTYGTTCIADENNETIEAQEAL 301

Qy   346 TWQEPHFHYVPRDQVLAGTRVSTFOCEATGNQPQAIPIWRBRSQNNLFSTYQPSQSSRF 405
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    302 AVQVPEFLPKAPANTIAESMDIYPECEVGTGPTPTYVWKNKGDIVIPSDI-----F 353

Qy   464 SVSQTGDLTIINVQRSDGYGIQCLLNAGSIGITIAKLEVTDIADRPF-PVPIRQGPNV 463
     ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db    354 KIVKENNLVGLGLWSDEQPTCYAENDVGNQAQAGAILLLDLDAIPLTPLTSLTSATN 413

Qy   464 QTAVAGDFFVLCVSATGSPVPTILRWKQDVLYSTQPSRIQLENGLVATRYAKLGDGTGR 523

```

CC or send an email to license@ish-sib.ch)

[illegible]

Db 1316 AVKMMKSGTSLVTDIGRRSIFNSGSPFIITVKARDSGTSCITANNMNSDEIILNLQ 1375
 Qy 1025 -----NQKTNMLPESTVYGD-----VDLSKNINEM-----KT 1052
 Db 1376 VQVPPQPRLVTSKTTSSITLSNLPGDNGSSIRVILYQSDNDSQWGSFPIPSERS 1435
 Qy 1053 FNSNMLDKGRF-----WNFS-----GQPTFYATTLQISNLNNMNN----- 1089
 Db 1436 YLENLKCTGWTWKTFLTAQNG/GPGRISELIRATLKGPOKSKQELFASINTIRVLN 1495
 Qy 1090 -----GSGDGS-----EKNKPLDQKQVAPQVQVINDKLNKT----- 1126
 Db 1496 LIGNDGQCPITISFLEYRPP-----TTWTATQRTSLSKSYLLDQLQATWYELQ 1548
 Qy 1127 -----RAN-----DTPVPTIPYNSQTDQNGSSYSSDGSSTSGSQGRK- 1166
 Db 1549 RVCNSAGCAEQANFATLNTDSTIPLI--KSVYQN-----EGLTT--NEGKLM 1595
 Qy 1167 -----KGARTPKVPK-QQGMWADLLPPFPHPPHNSSEYEN 1203
 Db 1596 LVTISICILVGVLLVLLVLRARRERQRLKDAKSLAELM-----MSKNTRTSD 1647
 Qy 1204 LSVDESDIQEMCPVPPARMYQQDELEEDERGR-----TPPVGAASSPAVSYSH----- 1257
 Db 1648 TLKSGQTLRWHDIPRAQLLIEROTMETDTSVLLTDADPGEAKQSLVTVTHVE 1707
 Qy 1258 -QSTAILPSPQELQPMQLQCPETGGMHQPDORRQPVSPPPP-----PRP 1304
 Db 1708 YQGSVGR-----GPIVDVSADARG-----TWFTTRNKAGATARNRYQWTLNRHP 1757
 Qy 1305 ISPHPTGYISGPLVSDMDTAP-----EEEDDADGVAKWRELLRGLQETPASS 1358
 Db 1758 TISARH-----LITDWRLLPFRAGSVKEDSDVSFSGQDTR-----ARSS 1800
 Qy 1359 VGDLE-SSVYSGMINGSASAEEDNISGCRSSVSSGSDSFTDADQAVAAATAGL 1416
 Db 1801 MYSTGASSTYELARAYERAKMEDQLRAKFTITE-----CFDSTSSQLTAGNEYDTS 1857
 Qy 1417 KVARQMDQAGRHFASQCPRTSPVSDTMSMAA--QKTRPAKLLKROPHGLRRTY 1476
 Db 1858 LITSFSE--SGICRTAS-----PPKPGQGRVNMVA/KAHRPG-DLMLPPLMADLF 1910
 Qy 1477 TDDLPPPPVPPPAIKSPITQ-SKTQLEVRVVPVKPLSMQARTDRSSQSGSSTK----- 1530
 Db 1911 LNRGQGTSDRLSLGQACLEPKSRITLKPFTVLEIP-MEASASSTREGSQWQGANVA 1969
 Qy 1531 ---GREVLGRQVDMRTNPGDPREAGQNDQNG--RGNKARLADLPAPKTHLI 1580
 Db 1970 TLPORGAGELQAAKMS-----SQESLLDSRHLKGN-----PYAKSTYL 2011

RESULT 6

NEOL_RAT
 ID NEOL_RAT STANDARD; PRT; 1377 AA.
 AC P97603;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEOGENIN PRECURSOR (FRAGMENT).
 GN NEOL or MGN.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=97015074; PubMed=8861902;
 RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-T.,
 RA Culotti J.G., Tessier-Lavigne M.;
 RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
 RL Cell 87:175-185(1996).
 CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE

CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
 CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
 CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
 CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
 CC TUMOR SUPPRESSOR PROTEIN DCC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENGL: E08726; A041100.1; -.
 DR HSPF: P56276; 1TLX.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFAM: PF00041; fn3; 6.
 DR PFAM: PF00047; lg; 4.
 KW Transmembrane; immunoglobulin domain; Glycoprotein; Signal.
 FT NM_187117.1
 FT SIGNAL 1 2 POTENTIAL.
 FT CHAIN 3 1377 NEOGENIN.
 FT DOMAIN 3 1074 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1075 1095 POTENTIAL.
 FT DOMAIN 1096 1377 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 36 105 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 135 197 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 232 296 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 324 386 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 405 502 FIBRONECTIN TYPE-III.
 FT DOMAIN 505 598 FIBRONECTIN TYPE-III.
 FT DOMAIN 599 698 FIBRONECTIN TYPE-III.
 FT DOMAIN 704 798 FIBRONECTIN TYPE-III.
 FT DOMAIN 819 919 FIBRONECTIN TYPE-III.
 FT DOMAIN 920 1021 FIBRONECTIN TYPE-III.
 FT DOMAIN 1087 1090 POLY-VAL.
 FT CARBOHYD 42 42 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 878 878 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1377 AA; 150637 MW; E514ED8ABD1A63A9 CRC54;

Query Match 8.64; Score 746.5; DB 1; Length 1377;
 Best Local Similarity 22.64; Pred. No. 3.8e-26;
 Matches 361; Conservative 213; Mismatches 636; Indels 385; Gaps 60;

Qy 67 PPRIVEHSDLVKSGEPATLNAKGRPPTTISWYKGERVETDNDPFRSHMLPSGS 126
 Db 21 PLTFYLVHSDLVKSGEPATLNAKGRPPTTISWYKGERVETDNDPFRSHMLPSGS 77
 Qy 127 LFLFVLVHGRSGPDGQYVYVAV-NYLGRVSHNASLEVALLDROFNDSVMVWAGE 185
 Db 78 LFLFVLVHGRSGPDGQYVYVAV-NYLGRVSHNASLEVALLDROFNDSVMVWAGE 136
 Qy 186 PAYMCKQPPRHPPTTISWYKSGDPLDKDERTITRSGKLMTITRSDAGKYVCTGVM 245
 Db 137 SGILNCE-VNADVPFVWQBNRQPLLDLIVKLPSGLVTSNATGDSGLRCLVIGES 195
 Qy 246 VGERSEVAVTLVLERPS-----FVKRPSNLAVTDGSAFCKEARGDPVPTVWRKDD 299
 Db 196 GPKPESDRAELKVLQSEEMLDLVLPLRPSMKIKVQGSVPLCVASGLPAPVWRKNE 255
 Qy 300 GEL-----PSKRIERDHTLIRKIVTAGDMSGYCVARNMVGAKASATLTVQEPHFV 356

Db 256 DVLDTSGSRLLAGGLAGSLEISDVTEDDAGTFCVANDGNKTEIAQALQLVQVPPFLLIQ 315
 Qy 357 PRDQVWALGTFTVQCEATGNPQPAIFWRREGSONLLFSTYOPPOSSRSFVSQGLDIT 416
 Db 316 PAMVARESDIMVFCEVTKRPAFTVKWKGNDVIPSQY-----FLVKEHLQVL 367
 Qy 417 NVQSGDWGTYICQTLNAGSILITRAKLEVTVDIADRPVPIRQGVNQVAVQGVFLSC 476
 Db 368 GLVKSDBGFTQCAENDVQNAQAQL-----ITLERAP----- 401
 Qy 477 VATSGPVPTILMRKQDVLSTQSDRIKLENGVLQIRYAKLGDTGRYTCIASTPSGE-AT 535
 Db 402 -ATGCTPLSPARDVWASLVST-----RFKL-----TWRTPASDHGMDLT 441
 Qy 536 WSAIVTEVQEFVQVQPRPTDPLNLPSSAKPEVTVSRNVTLSWQPNLMSGATPTSYI 595
 Db 442 YSVFTTKE--GVAREVENT-----SOPGEMQVI----- 468
 Qy 596 IZAFSHAGSSSQVQVQVAKNKTETSAIKGLKPNATYFLVRAANAYGISDPSQISDPVQ 655
 Db 469 -----IQNLMPATYIFKVAQNKNG-SSGSSAPLAVETQ 502
 Qy 656 DVLPTSGQVQHQVQRLGNVHLNHPVLLSSSLEVEN-TWQDSQITQYKILTPS 714
 Db 503 PEV-----QLGPAPIRATATISPT-----STVWETPLSGGEGQNKILITK 548
 Qy 715 GANHGSDMLVFVETPAKNSVVDPLARKGVTEIKARFPFNEQDASGEIKFALTEA 774
 Db 549 GTDK-EQDWDV-----SSHSYTLNGKITYETSFVAVTNKHPGVSTQDAVRLTSQ 601
 Qy 775 PSAPPGQVTVSKNGDGTALVSWOPPPEDTQNGHGVKWCMLGNTRHINKT-WDGS 833
 Db 602 PSAPPLNLSLEVR--NSKSIVLHWQPSAQTQNGQITGKIRYKASASVDVETVYGT 659
 Qy 834 TFSVYIPFLPGIRTSVEVAASTGAGSVKSE-----PQFILDHAGNP 877
 Db 660 QLSQILBGLDGTETNFRVALVNTGTPATDMLSAETPESDLOSBSVPE-VPSLAVRP 718
 Qy 878 V-----SPEDVSLAQISDVQKAPAFIAGACNWLIMVTSWLYRHHKKNGL 928
 Db 719 LVTSIVSMTPEPQ-----NIVRGYAGTIGSGPQAQITKD-----YKOR----- 761
 Qy 929 TSTYAGIRKV-PSFTPTPTV-TYORGEAV-----SSGGRP-----GLNINSE-ARQ 974
 Db 762 ---YTIENLDPSSSHYVILKAFNNWGEGLPLTESAVTPRHTDTEVDLVLINATPTVP 818
 Qy 975 WLADTVPWTG-----NNHDCSISCTAGNNGSMDLTYISRAPIANTNNQLNQTIN 1029
 Db 819 DPTPMPMPVGVQASILSDHTIRTW-----ADNLSLKHQITD--SRITYT--RMTN 867
 Qy 1030 LMLPESTYVQGVOLDNLNEMKMTFNSPN-----LKDGRVNPSPGQPTTATQLIQS 1081
 Db 868 ---IPANTKYKIAN-ATTLSYLVGLKPLMTLYEFSVMTVGRKRSWTSMAGHATFLVPT 924
 Qy 1082 NLSNNMNGSGSGSGEK-----HWKLGQOQEVAP-VQYVIEQKMLKLDIANDTVPPTI 1136
 Db 925 SPPDKPVSYSKSGKPKPTILVNWQPSSEANGKITYLIYSTVDNAEHDWVLEPVGRL 984
 Qy 1137 PFV-OSTYDQTTGGSYNSDGSSTSGSGHQHKGARTPKVPKQNGNNWADLPPPPA--- 1191
 Db 985 THQIQBLTDPPTFKIQRNKGMGSPSAVQFTPKADS-----SDKMPNDQALGSA 1038
 Qy 1192 -----HPPHSHSNSEYINVSDESYDQEMCPVPPA-----RMYLQ 1227
 Db 1039 GGGKGLDPLGSDTTPMSSGNSPHSGSTPLSDSNLIIIVSIVITVIVVVIIVFCTR 1098
 Qy 1228 DELEEDEREDGPTTPPVGAASSPAVSYSHQSTATLTPSQEOLQMLQDPEETHGHW 1287
 Db 1099 RTYSQKQKRAKXSVNG-----SHYKGNCKVDKPDMLT-----H 1135
 Qy 1288 QPDRRQVPSVPPPPFIPSPHTYGYISGLVSDMTDAPEEEDADEM-----VAKMQTR 1344

Db 1136 HERLEKPIDGSDPDNPMV-----TDTPIRANSQDITVWDSNMSNIHQ 1180
 Qy 1345 RLLRLAGLQTPASSVQDLESSVQSGMNGWSAEDNISGRSS--VSSSDGSFTDA 1401
 Db 1181 RNSTYHESDESMSITLAGRGRMKPMMPFDSQPPQSVNTPTSTDTMPASSSQCTCDD 1240
 Qy 1402 DFAQAQAAAYATAGLVARBMQDAAGRHHFASQCPRTSPVSTDSNMSAAMQKTRFA 1461
 Db 1241 QDPEG-ATSSSY-----LASSQBED-----SGSLPTAFW-----RPS 1272
 Qy 1462 KKLKHQPGHARETITDLPVPPVPP-----PAIKSPQASKTGLE----- 1502
 Db 1273 HPLK-----SFVAPAIPPGCPPIYDLPALPTPLLSQALNHLHSVKTSATSLGR 1323
 Qy 1503 ---VPPVVPVPLPSMDARTDRSSDGSVYKGEV 1534
 Db 1324 SRPMPVPPVPSAPVEQATRMLEDSE--SYPEDEL 1357

RESULT 7
 CAML_RAT
 ID CAML_RAT STANDARD; PRT; 1259 AA.
 AC Q0565;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN L1CAM OR CAML1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N. A.
 RX MEDLINE:91372414; PubMed:1894011;
 RA Miura M., Kobayashi M., Asou H., Uyemura K.;
 RT "Molecular cloning of cDNA encoding the rat neural cell adhesion
 RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced
 RT by differential splicing";
 RL FEBS Lett. 289:91-95(1991).
 CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO ANKONIN ON NEURONS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING.
 CC -1- TISSUE SPECIFICITY: THE SHORTER ISOFORM IS PREDOMINANTLY FOUND IN
 CC THE BRAIN, WHILE THE LONGER ISOFORM IS FOUND IN THE PERIPHERAL
 CC NERVOUS SYSTEM.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X59149; CAA41860.1; -
 DR PIR: S17655; S17655.
 DR HSP: P20241; ICFB.
 DR INTERPRO: IPR001777; -
 DR INTERPRO: IPR003006; -
 DR PFAM: PF00041; fn3; 4.
 DR PFAM: PF00047; ig; 6.
 DR PRINTS: PR00014; FNTYPE11.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT ANAL 1 19 BY SIMILARITY.
 FT CHAIN 20 1259 NEURAL CELL ADHESION MOLECULE L1.
 FT DOMAIN 20 1122 EXTRACELLULAR (POTENTIAL).

Qy 452 RPPPVIRGPGVNOT-VAVDG-TEVLSCVATGSPVPTILWRKDGVLVSTODSRIKOLENGV 509

```

RESULT 8
NEOL_HUMAN
ID NEOL_HUMAN STANDARD; PRT; 1461 AA.
AC Q92859; 000340;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEOL CR NGW.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE=97236553; PubMed=9121761;
RA Meyerhardt J.A.; Look A.T.; Bigner S.H.; Pearson E.R.;
RT "Identification and characterization of neogenin, a DCC-related
RT gene.";
RL Oncogene 14:1129-1136(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE=97312699; PubMed=9169140;
RA Vleisnetter J., Chen X.-M., Miskewich F., Lane R.P., Yakawaka K.

```

[illegible]

Query Match 8.3%; Score 725.5; DB 1; Length 1461;
Best Local Similarity 22.4%; Pred. No. 3.6e-25;

Qy 1084 SNNWNGSGDSEK-----HKKPLGQKQKV-----APQTNIVQ-----NKLAK 1124
 Db 958 KADVTYVSGAGPATIIVNQPSEANGKITGYITISTDWNAEIHWVIEPVNGLR- 1016
 Qy 1125 DFRANDTVPPTIPNQSDQNTQGG-----SYNSDDR--GSSTSGSGQKHKGA 1169
 Db 1017 -HQIQLTLDTLTPYTKIQRNSGKMGSEAVQFRTPKADSSDKMPNDQASGSG--KGS 1073
 Qy 1170 RTPKVPKQGGNADLPPPARPPSPNSSEYNSVDESTDQEMCPVPPA----- 1221
 Db 1074 RLPLD-----GSDYKPPMSGNSPHG-----SPISPLDSNMLLVIVSVGVITIV 1119
 Qy 1222 -----RMYLQDDEEEDERSPPPVPGGAASPAVSYSHQSTATLTPSPQELQPLQD 1277
 Db 1120 VVIIVFCFTRTTSQKKRAKCSVNG-----SHYTKGNSKDVKPOLM----- 1165
 1278 CPEOTGMQHQPRDRQPVSPPPRPISPPHYTIGSLVSDMTDAPREEDDADME 1337
 Db 1166 -----NHERLEKLPIDKDPNPM-----TDTPIFNSQD----- 1196
 Qy 1338 VAKQMYTRILLGLQETPASSGDLSESVTSGMNGWGSSEADNIS--SGRSSVSSGD 1395
 Db 1197 -----IPVWNSMDSNHEQRNSYRGH-----ESDSMSTLAGRGMRPKM 1238
 Qy 1396 SFTDADFAQAAVAAAYTAGLVARRQMDAAGRRHFASQCPRTS-----PWS 1445
 Db 1239 MFF-DSQPPQVTSANRHSID-----NPHHFHSSSLAGPARSHLYHPGSPWFG 1288
 Qy 1446 TDSNMS--AAVQKTRPAKLLHQPHLRITRETYD-----DL 1480
 Db 1289 TMSLSDRANSTSVNTSPSTQMPASSSQCTCDHQPDEGATSSSYLASSQEDSGSL 1348
 Qy 1481 P-----PPVPP-----PAIKSPQAQKTLQLE----- 1502
 Db 1349 PTAHVPSPLASFAVPAIPEPPTDPAISPTLLSQALNHHSVKTASIGTLGRS 1408
 Qy 1503 --VRPVPVVKPSMDATDRSDRKGSSYKREY 1534
 Db 1409 RPPMVPVVSFAFEVQ-ETTRMLDENSESSYPEDEL 1441

RESULT 9

NRCA_CHICK

ID NRCA_CHICK STANDARD; PRT; 1284 AA.

P35331;

01-FEB-1994 (Rel. 28, Created)

01-FEB-1994 (Rel. 28, Last sequence update)

15-JUL-1999 (Rel. 38, Last annotation update)

DE NG-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.

RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;

RX MEDLINE=91258407; PubMed=2045418;

RA Grumet M., Mauro V., Burjoon M.P., Edelman G.M., Cunningham B.A.;

RT "Structure of a new nervous system glycoprotein, Nr-CAM, and its

RT relationship to subgroups of neural cell adhesion molecules.;"

RL J. Cell Biol. 113:1399-1412(1991).

RN [2]

RP SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE-EMBRYONIC BRAIN, AND RETINA;

RX MEDLINE=91258410; PubMed=1512296;

RA Kayem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;

RT "Bravo/Nr-CAM is closely related to the cell adhesion molecules L1

RT and Nr-CAM and has a similar heterodimer structure.;"

RL J. Cell Biol. 118:1259-1270(1992).

CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN

CC NEURON-NEURON ADHESION; NEURITE FASCICULATION, OUTGROWTH OF

CC NEURITES, ETC. SPECIFICALLY INVOLVED IN THE DEVELOPMENT OF OPTIC

CC

CC FIBRES IN THE RETINA.

CC -1- SUBUNIT: HETERODIMER. COMPOSED OF AN ALPHA AND A BETA CHAIN.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS ARE PRODUCED BY

CC ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: RETINA AND DEVELOPING BRAIN.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING NEURAL RETINA AND

CC EMBRYONIC BRAIN TISSUE.

CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X58482; CNA1391.1; .

DR EMBL: L08960; AAA8632.1; .

DR HSSP: P20241; 1CFB.

DR INTERPRO: IPR001777; .

DR INTERPRO: IPR003006; .

DR PFAM: PF00041; fn3; 5.

DR PFAM: PF00047; ig; 6.

DR PRINTS: PRO0014; FNTYFEI1.

CV Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;

CV Transmembrane; Alternative splicing.

FT SIGNAL 1 24

FT CHAIN 25 1284

FT DOMAIN 25 1143 NG-CAM RELATED CELL ADHESION MOLECULE.

FT TRANSMEM 1144 1166 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 1167 1284 POTENTIAL.

FT DOMAIN 56 125 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 155 220 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 261 323 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 351 415 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 445 508 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 536 599 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 638 699 FIBRONECTIN TYPE-III.

FT DOMAIN 738 799 FIBRONECTIN TYPE-III.

FT DOMAIN 837 906 FIBRONECTIN TYPE-III.

FT DOMAIN 943 1006 FIBRONECTIN TYPE-III.

FT DOMAIN 1057 1114 FIBRONECTIN TYPE-III.

FT DISULFID 63 118 POTENTIAL.

FT DISULFID 162 213 POTENTIAL.

FT DISULFID 268 316 POTENTIAL.

FT DISULFID 358 408 POTENTIAL.

FT DISULFID 452 501 POTENTIAL.

FT DISULFID 543 592 POTENTIAL.

FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 581 581 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 595 595 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 834 834 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 969 969 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 995 995 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1059 1059 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPLIC 612 621 MISSING (IN ISOFORM AS10).

FT VARSPLIC 1027 1038 MISSING (IN ISOFORM AS12).

FT VARSPLIC 1039 1131 MISSING (IN ISOFORM AS93).

FT VARSPLIC 1202 1205 MISSING (IN ISOFORM AS-CT22).

RESULT 10
 DCC_MOUSE
 ID DCC_MOUSE STANDARD; PRT: 1447 AA.
 AC P70211;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-MAY-2000 (Rel. 39, Last annotation update)
 DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
 GN DCC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=BRAIN;
 RX MEDLINE=96112625; PubMed=8570174;
 RA Cooper H.M., Armes P., Britto J.J., Gad J., Wilks A.F.;
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer
 RT gene (mDCC) and its expression in the developing mouse embryo.";
 RL Oncogene 11:2243-2254(1995).
 RN [2]
 RP REVISIONS.
 RC STRAIN=BALB/C; TISSUE=BRAIN;
 RA Cooper H.M.;
 RL SUBMITTED (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
 CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
 CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
 CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
 CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION, HIGHER LEVELS
 CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
 CC AND REMAIN AT THIS LEVEL IN THE ADULT.
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
 CC OR SEND AN EMAIL TO license@isb-sib.ch)

CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELLS TYPES OF THE INTESTINE.
CC -1- DISEASE: COLORECTAL TUMORS THAT LOSE THEIR CAPACITY TO
CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNUSUALLY LACK DCC
CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

DR EMBL: X76132; CAA53735.1; -.
DR EMBL: M32292; AAA52151.1; -.
DR EMBL: M32286; AAA52174.1; -.
DR EMBL: M32288; AAA52175.1; ALT_SEQ.
DR EMBL: M32290; AAA52176.1; -.
DR EMBL: M63696; AAA52177.1; -.
DR EMBL: M63700; AAA52178.1; -.
DR EMBL: M63702; AAA52179.1; -.
DR EMBL: M63718; AAA52180.1; -.
DR EMBL: M63698; AAA52181.1; -.
DR PIR: A54100; A54100.
DR PIR: A40098; A40098.
DR PIR: A38442; A38442.
DR HSP: P56276; 1TLK.
DR MIM: 120470; -.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFM: PF00041; fn3; 6.
DR PFM: PF00047; lg; 4.
DR PRINTS: PR00014; FNTPEI11.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Anti-oncogene; Disease mutation; Polymorphism.
PT SIGNAL 1 25 POTENTIAL.
PT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC.
PT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
PT TRANSMEM 1098 1122 POTENTIAL.
PT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
PT DOMAIN 54 124 1G-LIKE C2-TYPE DOMAIN.
PT DOMAIN 154 219 1G-LIKE C2-TYPE DOMAIN.
PT DOMAIN 254 317 1G-LIKE C2-TYPE DOMAIN.
PT DOMAIN 345 407 1G-LIKE C2-TYPE DOMAIN.
PT DOMAIN 426 522 FIBRONECTIN TYPE-III.
PT DOMAIN 525 618 FIBRONECTIN TYPE-III.
PT DOMAIN 619 716 FIBRONECTIN TYPE-III.
PT DOMAIN 722 816 FIBRONECTIN TYPE-III.
PT DOMAIN 840- 940 FIBRONECTIN TYPE-III.
PT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
PT DISULFID 61 117 BY SIMILARITY.
PT DISULFID 161- 212 BY SIMILARITY.
PT DISULFID 261- 310 BY SIMILARITY.
PT DISULFID 352- 400 BY SIMILARITY.
PT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 299- 299 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 319- 318 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
PT VARIANT 168 168 M > T (IN OESOPHAGEAL CARCINOMA).
PT /FTIG-war_003909.
PT VARIANT 201- 201 R > G.
PT /FTIG-war_003910.
PT VARIANT 1375 1375 P > H (IN A COLORECTAL CARCINOMA).
PT /FTIG-war_003911.
PT CONFLICT 138 138 MISSING (IN REF. 3).
PT CONFLICT 233- 329 MISSING (IN REF. 3).

```



```

CC      GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC      -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch.)
CC
DR      EMBL: X58274; CAA48335.1; -.
DR      EMBL: X57734; CAA7963.1; -.
DR      PIR: S28830; S28830.
DR      MIM: 190197; -.
DR      INTERPRO: IPR001777; -.
DR      INTERPRO: IPR003006; -.
DR      PFAM: PF00041; fn3; 4.
DR      PFAM: PF00047; ig; 6.
KW      Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW      Cell adhesion; Repeat.
CC      SIGNAL      1      28
CC      FT CHAIN      29      1012      AXONIN-1.
CC      FT PROPEP     1013      1040      REMOVED IN MATURE FORM (POTENTIAL).
CC      FT DOMAIN     54      118      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     148      216      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     254      313      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     341      402      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     433      495      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     523      594      IG-LIKE C2-TYPE DOMAIN.
CC      FT DOMAIN     606      612      GLT/PRO-RICH.
CC      FT DOMAIN     611      706      FIBRONECTIN TYPE-III.
CC      FT DOMAIN     714      809      FIBRONECTIN TYPE-III.
CC      FT DOMAIN     816      908      FIBRONECTIN TYPE-III.
CC      FT DOMAIN     917      1003      FIBRONECTIN TYPE-III.
CC      FT SITE       794      796      CELL ATTACHMENT SITE (BY SIMILARITY).
CC      FT CARBOHYD    75      76      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    198      198      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    204      204      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    461      461      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    477      477      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    498      498      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    525      525      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    830      830      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    918      918      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT CARBOHYD    940      940      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT LIPID      1012      1012      GPI-ANCHOR (POTENTIAL).
CC      SO SEQUENCE   1040      1012      113393 MW: 25487803C28FE6B CRC64:

```

Query Match 7.5%; Score 653.5; DB 1; Length 1040;
Best Local Similarity 24.4%; Pred. No. 3.7e-22;
Matches 227; Conservative 115; Mismatches 400; Indels 187; Gaps 26;

Qy	63	QEDFPPIPHPSDLIV-----SKEGPAITLCKAGKGRPTPIIYVNGKGEVETDKDPSRSH	119
Db	32	QTTGVQPVEDVGLSVLFLPEESTEQVLLACARASAPATYTNMGNTM-----KLEPSSRH	68
Qy	120	MLLPSSGSLFLRILVHGKRSKPRQGVYVYCVARVILGAVSHNASLVAIIRDDQFNPQMSDV	179
Db	99	QLF-GGNL-----VIMNPIKADAGVYVGLASNPVGTIVSRSLALRPGVLFQBSKEERQPV	143
Qy	180	MVAVCEANVMECPPHGPEPIITSNKDGSP-----LDDKERIIRIIGKMLTIYTKSDAG	236
Db	144	KAEHGCGWMLCPNPPHYVGLSVRWLLNEFPNIPITDGRHFVQVOTGMILYIARTNASDG	203
Qy	237	KYVCVYTNMNGERESEV-----AEILVERPSVYKR-PSNLAVTDDSAEFKECA	285
Db	204	NYSLCATSHMDFTSKSVSPKQALNLAEDTLFAPSIKARPAETVALVQOQVTLCEPA	263
Qy	286	RGDPVITVYVKKRDKGLPKSYRETRDHTLILKRYVTDGMSITCYCAANMNGKASAL	345

```

Db 264 FGNPVPPKKRWKVGDSL - SPQWTAPEPTLQIPSVSEDEGTEHCEAENSAGRGTVQGR 321
QY 346 TWQEPHFVVPKQDVALGRVTFQCEAENGQOPALVFWREGCSNLLFSTYQSPSSFR 405
Db 322 IVQAQPEMLKVIJSDTADIGSNLWKCAGKPRPTVRLWRNG -----EPLASQNRV 373
QY 406 SVSQGTGLTINWQSDWQYICQPLAVASGITKAYLEVVDVIADRPVPPYKQGVNQT 465
Db 374 EV-LAGDLRFKSLSDSGMTCVQAEKNHGTYASAEALQALAFD -----FRNPVRL 431
QY 466 V-AVDGTFVLSCVARGSPVPTILNRGQVLVSTDSRIKQLWNLQYVIAKLGDPGY 523
Db 428 IPAARGGELLAPQAPRAAKVWLNSK-GTEILWNSSVTVTFPGCLTIIRWISSRDBGY 486
QY 524 TCAISTPSGSAWSAIIEWQFEGVQVPVPPRTDPNL ----- 559
Db 487 TCAENFMKGNSTGILSYDRTKITLPPSSADIMGNLTLQCHASHDPTMDLFTFTWL 546
QY 560 ----- 559
Db 547 DDDPFDIDFKPGHYRTNNKWTIGDLTILNAQRHGKGTICMAQTVVDSASKEATVLWG 606
QY 560 IPSASCKPEYTVSRNVTYLSWQNLNGSATPTSYIIEAFSHAGSSGNTVAENV ---- 614
Db 607 PGGPGLVVVDLIGDTITGLSRSGFNDSIPKATYLQARTPEAG-KMKQVRTIPANIEG 665
QY 615 KTESAIKGLPKMAYILEVLVRAANAYGISDPQISDPKVTQDLVQPSGVGHKQVQREL 674
Db 666 NAEYTAQGLGTPMDYETFVLIAHLGTGSPGSGSKIRTEAAPSVA ----- 713
QY 675 NAVLHLNPTVLSS -----SSIEVMTVDQSQYIQ -----GYKILYRPSGANGSGDWL 724
Db 714 -----PSLGGGGAGPELIVMT -PMSRYQNGDGFGYLLSPRQSGSTNQWT - 760
QY 725 VFEVTKPKNSVTV -----PDLRGVNYETKARFFNEFGADSEIKFATLEAPAPPQ 780
Db 761 -----ARVGDAGAQTVYNSGSRVPTPEVKIYSTRNGDGPESLITALVYSAEEPRVAPT 817
QY 781 GY-----TVSKDNGNTAILVSWQPPEDTQNGMVQETKV -WCLGNETHYIKNKTVDGSTF 835
Db 818 KWVAWGSGSSEM-----VTYSPVQD-MGILLGYETRTWKAGDEAADRVTRAGLDT 871
QY 836 SVVIFPLVGRITSVEVAASGAGSGYKS 864
Db 872 SARVSLGHPKTKTVTRAYNAGCTGAPAS 900

```

```

RESULT 13
AXOL_RAT
ID AXOL_RAT STANDARD; PRT; 1040 AA.
AC P22063;
DT 01-AUG-1991 (Rel. 19; Created)
DT 01-AUG-1991 (Rel. 19; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1).
GN TAXI.
CS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RX TISSUE=SPINAL CORD.
RC MEDLINE=90199890; PubMed=2317872;
RA Purley A.J., Morton S.B., Manalo D., Karagozian D., Dodd J.,
RA Jessell T.M.;
RT 'The axonal glycoprotein TAG-1 is an immunoglobulin superfamily
RT member with neurite outgrowth-promoting activity.';
RL Cell 61:157-170(1990).
CC
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -2- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -3- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS. AND IN ADULT

```

CC BRAIN, SPINAL CORD AND CEREBELLUM.
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED ON A SUBSET OF AXONS
 CC IN THE DEVELOPING RAT NERVOUS SYSTEM.
 CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M31725; AAA42201.1; .
 DR PIR: A34695; A34695.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFAM: PF00041; fn3; 4.
 DR PFAM: PF00047; fn3; 6.
 KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 KW Cell adhesion; Repeat.
 FT SIGNAL 1 30
 FT CHAIN 31 21015 AXONIN-1.
 FT PROPEP 21015 1040 REMOVED IN MATURE FORM.
 FT DOMAIN 56 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 218 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 315 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 343 404 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 435 497 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 525 596 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 608 614 GLY/PRO-RICH.
 FT DOMAIN 613 708 FIBRONECTIN TYPE-III.
 FT DOMAIN 716 811 FIBRONECTIN TYPE-III.
 FT DOMAIN 818 910 FIBRONECTIN TYPE-III.
 FT DOMAIN 911 1005 FIBRONECTIN TYPE-III.
 FT SITE 796 798 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 527 527 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 777 777 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 832 832 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1040 AA; 113042 MW; 6270726614CBAFB CRC64;

Query Match 7.54; Score 652; DB 1; Length 1040;
 Best Local Similarity 24.54; Pred. No. 4.3e-22;
 Matches 234; Conservative 117; Mismatches 428; Indels 176; Gaps 27;

Qy 52 NSLIGTYSRRLQDPPRIIVHPSDLIV---SKGEPAITLNCARGRPTPIENKGGKGV 108
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 23 SSSPQAGTPTATGTFIEQPIGLFPESASDQVTLACARASPATWKNMG--- 79
 Qy 109 ETKDQ-DPSSHMLLPGSSLLFLAIVHGRKSRPDGYCYVARNITGEAVSHMSLVAI 167
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 80 TDMLNPSGSRQLM-GGNL---VIMSPTTQDQVYCLASNPVGVTVSKVAIRLFG 133
 Qy 168 LADDQFNQSNVMWVAGVPAWMECPQPGHPEPTISWKKDGP---LOOKDERITIGKK 224
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 134 LQFSKSEEDPKVTHEGVMLECPNCPATYPLSTRWLLNEPNFPIIDGRHPSYQTGN 193
 Qy 225 LMITITTSKAGSYCVGTVMVGRSEEV-----AEITVLERPSFVKR-PSNIAV 273
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 194 LYIATNNSADLNGSLATSHMDSKTSVFSKQVCLNALAEPLFAPSTIKARFPPTTA 253
 Qy 274 TVDSDAFVCEARGDQVPTVWRKDGGLPSRYEIRDDTHLIRKVTAGIMSGSYTVAE 333
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 254 LVQGVTLRCFAPGNVPVRIKWRVGDGL--SPQWATEITQLPVSFEDGTGCEAB 311
 Qy 334 NNWGAESAALVTLQEPHPVVKPVQVVALGRVITFQCEATGNPQAPLNRREGSONL 393
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 312 NSKGTQVIGRIIVQAPQEWLIVSDTADIGSNLRWGCAAGGPRPVWRLANG----- 366
 Qy 394 FSTQPPQSSSSFSVSTQGLTITNVRSDDGYVIGQTLNVAIGSIITKALEVTVADIAPR 453
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 367 ---EPASLQNVFVY-LAGDLRSLKSLSDSGMCCQVAENKHGITYASAEALQALPAQD- 420
 Qy 454 PPVIRGGVNVQTV--AVDGTFLVSCVATGSPVPTILWRKGVVLSTQSRIRKQLVQLQ 511
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 421 ---FRNPVRLIPARGGEISILCPRAAPKATILWSK-GTEILGNSTRTVITSDGLI 476
 Qy 512 IRYAKLDTGRTGRTCIASPTSGEATWSAIVIEQVGGVPPVPPRPPTDNL----- 559
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 477 IRNISRSDGRTCTCAENFMGRANSTGLSVDRATKITLAFSSADINVGDNLTQLCHASH 536
 Qy 560 ----- 559
 Db 537 DPTNDLFTWTLDDFPIDFKPGHTRAKASKITGLTILNAHVHGGKQITCMAQTVWD 596
 Qy 560 -----IPASPSKPEVTVDSRANTVLSWQNLNNGATPTTIIIEAFSHASGSS 607
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 597 GTSKEATVLAGPPGPGGVVVRDIDGTTVLQSWSGRGNHSPKIATQLQARTPPSG-KW 655
 Qy 608 TVQAVN---KETSALGLKPNALYFLVRAANAYIGSDPSQISDPVKTQVLPVTS 662
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 656 KQVTRTFWVWIGNMETAQVGLMFMMDYFPRVSVASNLITGEPGSGPSKIRTKAVPSVA 715
 Qy 663 GVHKGQVQBELGNALVHLNPNVLS---SSIEVNT-VQQSQITGYKILIRPSGA 716
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 716 -----PSGLSGGGGAPGELLNWTVPVSEYQMGDGP--TLLSFR 753
 Qy 717 NHGESDNLVEFVTPAKNSVTL---PDLRGVNYEIKARPPFFHQDAGSEIKFATLE 772
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 754 RQSSSMQT--ARVPGADQYFVYVNDGSIQPTTPPVKIRSYNRGCDGPESUTALVSAE 811
 Qy 773 EAPSAQPGGVYVSKNDGNTALVSWQPPEDQNGVQVYV--WCLNGETRYHINKTV 830
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 812 HEPRVAP--AKWAKGSSSEDMVSWEPVLD-MNGLILGTETRYWAGDNEAADRVRT 868
 Qy 831 DGSFSPVIVPLVPGIRYSEVAESTAGGSGVKSQEPQLDQARNVPSPEQVDS 885
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 Db 869 AGLDTSARVTLGNWTKTHTVTRVNAKRTGPAS-PSADMTVTPPPRRPGNIS 922

RESULT 14
 AXOI_CHECK
 ID AXOI_CHECK STANDARD; PRT; 1036 AA.
 AC P2685;
 DT 01-DEC-1992 (rel. 24, Created)
 DT 01-DEC-1992 (rel. 24, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 OS AXONIN-1 PRECURSOR.
 DE Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 KC TISSUE-BRAIN:
 RX MEDLINE=92174893; PubMed=1311675;
 RA Zuellig R.A., Rader C., Schroeder A., Kalousek M.B.,
 RA von Bohlen Und Halbach F., Osterwalder T., Inan C., Stoeckli E.T.,
 RA Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
 RT "The axonally secreted cell adhesion molecule, axonin-1. Primary
 RT structure, immunoglobulin-like and fibronectin-type-III-like domains
 RT and glycosyl-phosphatidylinositol anchorage.";
 RL Eur. J. Biochem. 204:453-463(1992).
 CC -1- FUNCTION: AXON-ASSOCIATED CELL ADHESION MOLECULE (AXCAM) WHICH
 CC PROMOTES NEURITE OUTGROWTH BY INTERACTION WITH THE AXCAM L1 (G4)
 CC OF NEURITIC MEMBRANE.

Qy	66	PPPRIVFSDLSYSG--SPALNCKAGSPPTPIEWKGESEVETDKDPSRHML	122
Db	30	YGVVFEQPAHLFPGSASEKVLTKCARANPPTATWNNKNTLKMGPDS-----RRL	85
Qy	123	PSGLFFLRILVHGKSPDPBGVTCVARYLTGEAVSHASLEVALRDLQNPQSDVWA	182
Db	86	VAGD-----VISNPMKADGSCYCVANWARGTVVSRASLRPGFLFGSAERDPVKIT	141
Qy	183	VGEPVAMRGPPHGPPEPTISWKKGSP-----LDDKRLIRIIGKKMLITTYKSDAGKV	239
Db	142	EGWGMFTCSPPHYPALSTFWLLNEPFPPIADRRGVSQTTGMNLYAKTVEADLNS	201
Qy	240	CVGTMMVG-----ERESEVRLTVLERPS-VKRPNSLAVTVDSAEFKARGD	288
Db	202	CFATSHIDFTIKSVSKSQLSLAARDQAQVPSIKAFPADYALQGMWTLKCFPAGN	261
Qy	289	PVPTVWRKDGDELPSRVEYRDHDTLKRKTWAGMGSTYCVASNMVKGAEASLTIVQ	348

Db	262	PVQIKSNWKLDS-----QTSKWLSEPLLHIJNVDFEDETPECEAENKAGRDYTGRIIIL	319
QY	349	EPFPHVLPDQWALGRVITPQCAENGNQPAIFWREGSQNLLSTPQPPSSRSFVS	408
Db	320	AQWDLWDITTEADTIGDLKMSVCAGKRPRAWKLDRG-----QPLASQNRIVS	371
QY	409	QTGDLIIINWQSVQSYVCIQLAVASITIKAYLEVDIADRPVPPVIRQGVQWV	466
Db	372	-GGELFRSKLVLEDGQTCVAENKHGVSASAEUVALQAPD-----FLNPKWKLPIA	425
QY	467	AVDGTFWLSCVATGSVPVPIIWRKRDGLVYSTQDSRIKQLENGVLRIKATGDRGYTC	526
Db	426	ARSGVITPQDQAPRAAFKATVLTWK-GETLLNSGRVITADGTLILQNKISDEBGTTCF	484
QY	527	ASTPSGEATWSATIEVQEGVQVPPRPPTDML-----	559
Db	485	AENFMAKSTGLISLVADATITLAPASDINGENGLILQCHASHDPTMDLITFWNSFD	544
QY	560	-----	562
Db	545	PIDLDKSEGHTRASVKEAVGLDVLINVAQLKSGRTCTTCAQTVDSSTESATLVRKPGP	604
QY	563	APSKPEVTVDSRNTVITLSWQNLMSGATPTSYIIIEAFSAHGSSQWTVAEVY-----KTE	617
Db	605	PPGVVITDIDGTIVQLWSRSGPDHSPFIARSIBARTLLS-NKWKQMTNFPVNIAGNE	663
QY	618	TSAIKGLKPNATILFLVRAANAYISDPSSISDPVKTQDVLPTQSGVDHKVQRELGNV	677
Db	664	TAQVVLTPMDTFFFLVLANSLTGVEPSPSPSKRTKEAPVA-----	708
QY	678	LHLNPTVL-----SSSEVHWTV-----DQSQSYIQYTKILYSPSGANHGSDMLFVFR	729
Db	709	-----PSGLOGGGGAPNELIINWPTLRDQGDGFGILFSFRKGT-----QGLMT-AR	757
QY	730	TPAKNSVVIPLDLAGVY-----YEIKARPFNEPQAGDESIKFAKTLEEAPSAPPQGVTS	785
Db	758	VPAASSLATVITYRNESIGPTTFEVIKAYNRKGEPESLTAIVTSAEEKPKVAPFVIAK	817
QY	786	KNDGNGTALL-----VSWQPPEDDTQNGVQEKYV-WCLNGETHRYTHNKYVDSTFSV	838
Db	818	-----AVLSESDVSWPEPQSDMGQVILGLYEYIRWKDGKEAADRVATGLVYSAH	870
QY	839	IPFLVPGIRTSVEAASTAGSGVSKSEPOQTLDLDAHNGVSPSEDQVS	885
Db	871	VIGLNPNTKYIEVSRYANRAGG-----PPSPSTNIT	902

```

RESULT 15
NRG_DROME
ID NRG_DROME STANDARD; PRT; 1239 AA.
AC P20241; Q24414;
DT 01-FEB-1991 (rel. 17, Created)
DT 01-FEB-1991 (rel. 17, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE NEUROGLIAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RX MEDLINE=90030418; PubMed=2805067;
RA Bleher A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglial: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT Ll."
RL Cell 59:447-460(1989).
RN
[2]
RP SEQUENCE OF 1187-1239 FROM N.A.
RX MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bleher A.J., Patel N.H., Goodman C.S.;

```


RT "Differential splicing generates a nervous system-specific form of
 RT Drosophila neuroglian.";
 RL Neuron 4:697-709(1990).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
 RX MEDLINE=94213741; PubMed=7512815;
 RA Huber A.H., Wang Y.-M.E., Bleher A.J., Bjorkman P.J.;
 RT "Crystal structure of tandem type III fibronectin domains from
 RT Drosophila neuroglian at 2.0 A.";
 RL Neuron 12:717-731(1994).
 CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL
 CC ADHESION IN THE DEVELOPING DROSOPHILA EMBryo.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: NEURONS AND GLIA IN THE DEVELOPING NERVOUS
 CC SYSTEM AND ON SOME OTHER NONNEURONAL TISSUES.
 CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: M28231; AAA28728.1; ALT_SEQ.
 DR EMBL: X76243; CA53822.1; ..
 DR PIR: A32579; A32579.
 DR PDB: 1CFB; 3C-NOV-94.
 DR FLYBASE: FBN0002968; Nrg.
 DR INTERPRO: IPR001777; ..
 DR INTERPRO: IPR003006; ..
 DR PFAM: PF00041; fn3; ..
 DR PFAM: PF00047; ig; 6.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; 3D-structure;
 KW Immunoglobulin domain; Signal; Embryo; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 1239 NEUROGLIAN.
 FT DOMAIN 24 1138 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1139 1154 POTENTIAL.
 FT DOMAIN 1155 1239 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 53 123 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 149 124 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 262 329 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 354 422 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 447 512 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 536 606 1G-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 629 690 FIBRONECTIN TYPE-III.
 FT DOMAIN 729 792 FIBRONECTIN TYPE-III.
 FT DOMAIN 832 896 FIBRONECTIN TYPE-III.
 FT DOMAIN 932 997 FIBRONECTIN TYPE-III.
 FT DOMAIN 1024 1098 FIBRONECTIN TYPE-III.
 FT DISULFID 59 111 POTENTIAL.
 FT DISULFID 625 706
 FT CARBOHYD 162 182 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 821 821 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1234 1234 T -> Y (IN REF. 2).
 FT CONFLICT 1237 1237 L -> K (IN REF. 2).
 SQ SEQUENCE 1239 AA; 138284 MW; 49E12692D0DD02D7 CRC64;

QY 58 GSRLRQEDPPFPRIVEPS--DLIVSKGE-----PATLNKAEGRPTIDWYKGSERV 108
 Db 19 GSAESGNRRPRTIQPAPGELLFVKVQAQNKESDNPFIETCDADQPEPETSWMKNGKE 78
 QY 109 ETDKDDPSHSMPLSPGSLFLIRIVHGRKSRPDSGVYTCVARNLGEAVSHASLEVAL 168
 Db 79 DNQATNENMLKIQ--PGRGVLWITI-----PKDEDRGVYCFASNEFGTATNSVYKRAEL 132
 QY 169 RDOFPQNDPQWVAV-GEPAVMECPQPRGHPETPS----- 204
 Db 133 -NAFKDEAKLTLEAVEGEFPMKCAAPDGFSPPTVMWIMQESDGSIKNSNNSTLMDPE 191
 QY 205 -----KKD----- 207
 Db 192 GNLWFSNVNREDASSDPTTACSATSVRSXKIGNKVLVDVQWQVSAQNKHPVVRQV 251
 QY 208 -----GSPID-----DKD-----ERIT--IRGKLIMITITKASD 234
 Db 252 SRQSLALAGKRMELFCIYGTPPLQTVKSDGQRQSDAITQGHGKSLVIRQTNFDD 311
 QY 235 AGKTVCGTNMNGVERSEVALTVLERPSFKVPSNLAVTVDDSAEFKCRAGDPVPTVR 294
 Db 312 AGITITCDVSGWGNQSFSLINLVNVSPPYTPKEPEATAAEDEVVFECAAGVPEPKIS 371
 QY 295 WRKDDGELKSPYVEIR-----DDHTLAKRVTAGDMGTYTCVAENMGKASAILQV-EP 350
 Db 372 WIRNKGRLEQSTPNPRVITVDTNIRILNLVKGDTGNGCNATSLGVTYKDVYLWQAP 431
 QY 351 PHFYVKBPDQVVALGRVTVPQCEATGNPQALFVNRREGSNQLLSTYPPQSSRFSSVQT 410
 Db 432 PTISEAPAAVSTVDGRNTIKCRVNGSPKPLVKMLR--ASWML-----TGRVNVQAM 482
 QY 411 GDLTINLVQSGWGYIQTIAL-----VAGSIITKALEVTDVIADRPFPVYQVQV 464
 Db 483 GDLTITQVTSADGKTTCAQNKPFGEIDQSSVLVKEHTRIT-----QEPQNT 530
 QY 465 TVANDGTFLVSLCVATGSPVPIIL--NRKDGVLV--STQDSRIKLENGVLQIRKALGOT 520
 Db 531 EVAAGQSATPCRNKAHDDLTLEIDWNRDQSDIFPCAIPYKATNDNSLTIATKHEL-DE 589
 QY 521 GRITCIASPTSGEATSAIIEVQGEVFPVPPRTPMLPLSAPSKPEVTDVSRMIVLS 580
 Db 590 GECTVCARTLDEATARANLVQD-----VPAKALTGIT-CQAKAEIH 633
 QY 581 WQPNLNGSATPTSIITIAFSSHASGSGWQVAVNR-KTETSAIKGLKPAIILFVRAANA 639
 Db 634 WEQQQNSPILHTYITQFTSEPTASWDAAYEVQVNTDSSFVQNGSWANITFRVIAFNK 693
 QY 640 YGISDPSQISDPVKTVQLP-----TSGQVD-----HKQV-- 669
 Db 694 IGAPSPASBSGCTIQDQPPFKNPQNVNGVGTETPNLWISWTHPEIENNAFNFYTSV 753
 QY 670 QRELGNVLHLN----- 682
 Db 754 KRDIIPAAWNNKIPDWRQNNIVADQPTFKYLKLVAVINDRGSNVAEEVVGTSGED 813
 QY 683 -----PT-----VLSSSIEVWNT-VQOQS-QYIQGYLLRPSGANGSDWLVFV 728
 Db 814 RPLDAPNFTMQITISSTSGYMAWPVSEBSVGHGKGTIK--QTVENRGEGBELRNV 871
 QY 729 RPKANSVYI-----POLRKNVTEIKARPFNFEGQADSEIKFAKLEAPAPQGVTV 785
 Db 872 KGOTNALVTQFKPDSK--NY-ARILAYNGRNGPPSAVIDPTTSGVGS-PVQGL-- 923
 QY 786 KNDG--NGTALYSWQPP--PEDTQNGVQBYKVV-----CLGNETRY--HINKTV 831
 Db 924 --DAYPLSSAPMLNKKPLP-----NKLKLGKITYIEVKSEYVGERREYDPIH--TO 974
 QY 832 GSTFSWILPVLPGVIRTSVEVAAGTAGGS----- 861
 Db 975 PRVTMMKGLKPNRSKRISITATTKMGREGSEYIEKTLKADAVNAPATPFSWEOLES 1034

Query Match 7.2%; Score 627; DB 1; Length 1239;
 Best Local Similarity 22.0%; Pred. No. 7e-21;
 Matches 281; Conservative 148; Mismatches 427; Indels 420; Gaps 49;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:53:03 ; Search time 559.88 Seconds
(without alignments)
345.628 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKNKRVFPLVMSLLSLSPN.....VLGSTERGEDNNNELEETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Aligned: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organella.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	No.	Score	Match	Length	DB	ID	Description
1	8724	100.0	1651	4	Q9Y6N7			Q9Y6N7 homo sapien
2	8315	95.3	1651	11	O55005			O55005 rattus norv
3	8120	93.1	1612	11	O89026			O89026 mus musculus
4	3104	35.6	1060	11	Q9QZ13			Q9QZ13 rattus norv
5	2607.5	29.9	1344	11	Q9Z2T4			Q9Z2T4 mus musculus
6	1594	18.3	1395	5	Q9W213			Q9W213 drosophila
7	1592	18.2	1395	5	O44924			O44924 drosophila
8	1515	17.4	1273	5	O44928			O44928 caenorhabdi
9	1307.5	15.0	859	5	Q9VZ26			Q9VZ26 drosophila
10	1212	13.9	823	5	Q9VQ10			Q9VQ10 drosophila
11	873	10.0	285	4	O43608			O43608 homo sapien
12	852	9.8	423	5	P91572			P91572 caenorhabdi
13	760	8.7	1493	11	P97798			P97798 mus musculus
14	758	8.7	1259	11	Q9QY38			Q9QY38 mus musculus
15	757.5	8.7	1443	13	Q90610			Q90610 gallus gall
16	756	8.7	1822	4	Q9ULZ7			Q9ULZ7 homo sapien
17	748	8.6	1248	6	Q9XT41			Q9XT41 cercopithe
18	746.5	8.6	1377	11	P97603			P97603 rattus norv
19	726.5	8.3	1427	13	Q91562			Q91562 xenopus lae

20	725.5	8.3	1461	4	Q92859			Q92859 homo sapien
21	725.5	8.3	1461	4	O00340			O00340 homo sapien
22	718.5	8.2	1100	4	O94779			O94779 homo sapien
23	713	8.2	1026	4	O94780			O94780 homo sapien
24	707.5	8.1	1308	4	Q9UHI4			Q9UHI4 homo sapien
25	706.5	8.1	1026	11	Q62845			Q62845 rattus norv
26	704.5	8.1	1180	4	O15051			O15051 homo sapien
27	702.5	8.1	1028	11	Q62682			Q62682 rattus norv
28	702	8.0	1236	4	Q9UHI3			Q9UHI3 homo sapien
29	700.5	8.0	1299	4	Q92823			Q92823 homo sapien
30	696.5	8.0	2221	5	Q9UIM1			Q9UIM1 drosophila
31	695.5	8.0	1028	11	Q07409			Q07409 mus musculus
32	689.5	7.9	1099	11	P97527			P97527 rattus norv
33	689.5	7.9	1166	11	Q9QVNA			Q9QVNA rattus sp.
34	689.5	7.9	2016	5	Q9NBA1			Q9NBA1 drosophila
35	684.5	7.8	874	5	O01632			O01632 caenorhabdi
36	684.5	7.8	2016	5	Q9V4J9			Q9V4J9 drosophila
37	683.5	7.8	1445	11	Q63155			Q63155 rattus norv
38	681	7.8	1299	4	O15179			O15179 homo sapien
39	679.5	7.8	1215	11	P97686			P97686 rattus norv
40	678.5	7.8	2222	5	O97394			O97394 drosophila
41	674	7.7	1154	11	Q9QVX3			Q9QVX3 rattus sp.
42	666	7.6	1028	11	P97528			P97528 rattus norv
43	664.5	7.6	1028	4	Q9Q052			Q9Q052 homo sapien
44	657.5	7.5	1224	4	O00533			O00533 homo sapien
45	653	7.5	1028	11	Q9JMB8			Q9JMB8 mus musculus

ALIGNMENTS

RESULT 1
Q9Y6N7
ID Q9Y6N7 PRELIMINARY; PRI: 1651 AA.
AC Q9Y6N7.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBOT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors."
RT Cell 92:205-215(1998).
DR ENBL: AF040990; AAC39575.1; .
DR HSP: P56276; ITLK.
DR INTERPRO: IPR001777; .
DR INTERPRO: IPR003006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; lg; 5.
SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 100.0%; Score 8724; DB 4; Length 1651;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKNKRVFPLVMSLLSLSPNHLFLAQLIPDPEDVERGNHGHTPIPTSDNDNLSLGTGSR 60
|||||
Db 1 MKNKRVFPLVMSLLSLSPNHLFLAQLIPDPEDVERGNHGHTPIPTSDNDNLSLGTGSR 60
QY 61 LQEDFPFRIVFHPSDLVSKGEPATLNCKAEGRPPTTLEWKGGERVDKDDPSRSHW 120
|||||
Db 61 LQEDFPFRIVFHPSDLVSKGEPATLNCKAEGRPPTTLEWKGGERVDKDDPSRSHW 120

Qy 121 LLLPSGLFPLRIVHGRKSRPDEGVYVCARNYLGAEVSNHSALEVALRDFQNPQSDM 180
 Db 121 LLLPSGLFPLRIVHGRKSRPDEGVYVCARNYLGAEVSNHSALEVALRDFQNPQSDM 180
 Qy 181 VANGEPAYMEQPPRGHPEPTISWKDGSPLDDKDERITIRGKLMITTTKSDAGKYV 240
 Db 181 VANGEPAYMEQPPRGHPEPTISWKDGSPLDDKDERITIRGKLMITTTKSDAGKYV 240
 Qy 241 VGTVMHGERSEVAELTVLERPSFVKRPSNLAVTDSAEFKCEARGDPVPTVMKDDG 300
 Db 241 VGTVMHGERSEVAELTVLERPSFVKRPSNLAVTDSAEFKCEARGDPVPTVMKDDG 300
 Qy 301 ELKASRYEIRDOHTLAKRVYAGDMGSTCVAEWVNGAASALTQVQEPHFVPMKPDQ 360
 Db 301 ELKASRYEIRDOHTLAKRVYAGDMGSTCVAEWVNGAASALTQVQEPHFVPMKPDQ 360
 Qy 361 VVGLGRVTFQCEATGNPQPAIFWRKRGSONLLESTFQSSRSFVSQGDLTINWQ 420
 Db 361 VVGLGRVTFQCEATGNPQPAIFWRKRGSONLLESTFQSSRSFVSQGDLTINWQ 420
 Qy 421 SDWGTICQTLNVAISITKALEVTVIADRPVPIRGVWQVAVDGTFLSCVATG 480
 Db 421 SDWGTICQTLNVAISITKALEVTVIADRPVPIRGVWQVAVDGTFLSCVATG 480
 Qy 481 SPYPTILWMDGVLVSTQDSRIKQLENGVQIRAKLDGTGRTCIASPTSGEATWSAI 540
 Db 481 SPYPTILWMDGVLVSTQDSRIKQLENGVQIRAKLDGTGRTCIASPTSGEATWSAI 540
 Qy 541 EYQEFQVPPPPPTDNLPIAPSKEPVTIVSRNVLWSQWNLNSGATPTSTIIRAFS 600
 Db 541 EYQEFQVPPPPPTDNLPIAPSKEPVTIVSRNVLWSQWNLNSGATPTSTIIRAFS 600
 Qy 601 HASSGSQVIAENVKETSAISGLKPNATYLEVRANAYGISDPSQISDVPKTVQVLT 660
 Db 601 HASSGSQVIAENVKETSAISGLKPNATYLEVRANAYGISDPSQISDVPKTVQVLT 660
 Qy 661 SQGVDBHQVQRELGNVHLNPNPTVLSSEIEVHWTDQSQSYIGYKILRPSGANHG 720
 Db 661 SQGVDBHQVQRELGNVHLNPNPTVLSSEIEVHWTDQSQSYIGYKILRPSGANHG 720
 Qy 721 SDWLVEFVRTPAKNSVVIPLRKGVNYEIKARFPNEFGQADSEIKFATLEAPASPP 780
 Db 721 SDWLVEFVRTPAKNSVVIPLRKGVNYEIKARFPNEFGQADSEIKFATLEAPASPP 780
 Qy 781 GVTYSKNDGNGTALLVSWQPPEDTQNGWQYQYKVKLGNETRYHINKYDGSFVSYP 840
 Db 781 GVTYSKNDGNGTALLVSWQPPEDTQNGWQYQYKVKLGNETRYHINKYDGSFVSYP 840
 Qy 841 FLVPGIRYSYVAASTAGSGVSKSEPFQILDAGNVPSPEDQSLAQISDVWKQAPI 900
 Db 841 FLVPGIRYSYVAASTAGSGVSKSEPFQILDAGNVPSPEDQSLAQISDVWKQAPI 900
 Qy 901 AGIGAACWILMVFSIWLRYHRRKKRGLISTYAGIRKVSPTFTPTVYQGGEASVSG 960
 Db 901 AGIGAACWILMVFSIWLRYHRRKKRGLISTYAGIRKVSPTFTPTVYQGGEASVSG 960
 Qy 961 RGPLLINSIPAQWILADWPNWGNHNDCSISCTAGNNGSNDLITTSRPACIANY 1020
 Db 961 RGPLLINSIPAQWILADWPNWGNHNDCSISCTAGNNGSNDLITTSRPACIANY 1020
 Qy 1021 NQLNQLQNTMLPESTVYGVODVLSKNINEMKTFNSPLNADGRFVSGQPTTATTLQ 1080
 Db 1021 NQLNQLQNTMLPESTVYGVODVLSKNINEMKTFNSPLNADGRFVSGQPTTATTLQ 1080
 Qy 1081 SNLSNLMNNGSGSGSEKHWKPLGQOQEVAPQVYNIVEDNKLNDYRANDVVPPTIP 1140
 Db 1081 SNLSNLMNNGSGSGSEKHWKPLGQOQEVAPQVYNIVEDNKLNDYRANDVVPPTIP 1140
 Qy 1141 SYDQNTGGSYNSDSRGSSTSGSQGHKKGARTPVKVPQGNWADLPPPHAPPPHNSNE 1200
 Db 1141 SYDQNTGGSYNSDSRGSSTSGSQGHKKGARTPVKVPQGNWADLPPPHAPPPHNSNE 1200
 Qy 1201 EYINISVDESYDQEMCPVPPARMYLQDELEEBEDRGPTPPVKGAASSPAVYSHQST 1260

Db 1201 EYINISVDESYDQEMCPVPPARMYLQDELEEBEDRGPTPPVKGAASSPAVYSHQST 1260
 Qy 1261 ATLTPSQEELQPLMQDCPEETHGHQHPDRRRQVSPPPPPRISPPHYTIGISGLVS 1320
 Db 1261 ATLTPSQEELQPLMQDCPEETHGHQHPDRRRQVSPPPPPRISPPHYTIGISGLVS 1320
 Qy 1321 DMDTDAPEEBEDADMEVAKMOTRLLRLGLQEPASSVGDLESSTVSGMNGWSASE 1380
 Db 1321 DMDTDAPEEBEDADMEVAKMOTRLLRLGLQEPASSVGDLESSTVSGMNGWSASE 1380
 Qy 1381 DNISSGSSVSSSDGSFTDADFAQVAAAETAGLVARRQMDAAGRHFHASQCPRP 1440
 Db 1381 DNISSGSSVSSSDGSFTDADFAQVAAAETAGLVARRQMDAAGRHFHASQCPRP 1440
 Qy 1441 TSPVSTDSMSAAVMQKTRPAKILKQGHRLARETTDOLPPPPVPAIKSPITAGSTQ 1500
 Db 1441 TSPVSTDSMSAAVMQKTRPAKILKQGHRLARETTDOLPPPPVPAIKSPITAGSTQ 1500
 Qy 1501 LEVRPVVVVILPMDARTDRSSDRAGSSYKREVLDRGVWVMKTNFGDPEAQBDQMD 1560
 Db 1501 LEVRPVVVVILPMDARTDRSSDRAGSSYKREVLDRGVWVMKTNFGDPEAQBDQMD 1560
 Qy 1561 NGRGNKAARULPAPARTLHQEDILPYCRPTPTSMNPRDSSSSMSRSGSGSRQEDA 1620
 Db 1561 NGRGNKAARULPAPARTLHQEDILPYCRPTPTSMNPRDSSSSMSRSGSGSRQEDA 1620
 Qy 1621 NVGRNIAEMVLGGERGEDNNEELETES 1651
 Db 1621 NVGRNIAEMVLGGERGEDNNEELETES 1651

RESULT 2

055005

ID 055005 PRELIMINARY; PRT; 1651 AA.

AC 055005;

DT 01-JUN-1998 (TrnMBLrel. 06, Created)

DT 01-JUN-1998 (TrnMBLrel. 06, Last sequence update)

DT 01-OCT-2000 (TrnMBLrel. 15, Last annotation update)

DE TRANSMEMBRANE RECEPTOR ROBO1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID:10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SPINAL CORD;

RX MEDLINE=98117249; PubMed=9458045;

RA Kidd T., Brose K., Mitchell K.J., Petter R.D., Tessier-Lavigne M.,

Goodman C.S., Tear G.;

RT "Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors."

RL Cell 92:205-215(1998).

DR EMBL: AF041082; AAC39960.1; -.

DR HSP: P56276; 17LK.

DR INTERPRO: IPRO01777; -.

DR INTERPRO: IPRO03006; -.

DR PFAM: PF00041; fn3; 3.

DR PFAM: PF00047; 4g; 5.

KW Transmembrane.

SQ SEQUENCE 1651 AA; 180746 MW; FA2452D46E18687 CRC64;

Query Match

95.3%; Score 8315; DB 11; Length 1651;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 1560; Conservative 42; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKNKHPPVTVLISLLSPNHLFLAQLPDPEDVERGNHGTPTPPSDNDNSLGTGSR 60

Db 1 MKNKHLLPLVNLISLLTKSHLLLAQLPDPEDLERGNONGTPTPSDNDNSLGTGSR 60

Qy 61 LRQDPPPRIVRHPSDLVSKGEPATLNCAGRPPTTIEWKYGGERVETDQDPSRHW 120

Db 61 LRQDEPPPRIVHSDLTIVSGEPATLNCAGRPPTTIEWYKGERVETDKDQPSHRM 120

Qy 121 LLLSGSLFFLRIVHGKSRPDGSGYTCVARNYLGEAVSHNASLEVALLRDQFNPQSDWM 180

Db 121 LLLSGSLFFLRIVHGKSRPDGSGYTCVARNYLGEAVSHNASLEVALLRDQFNPQSDWM 180

Qy 181 VAVGEPAMVQCPGPRGHEPTISWKDQSPDLDDKERITIRGKMLITTTKSGDAKTYC 240

Db 181 VAVGEPAMVQCPGPRGHEPTISWKDQSPDLDDKERITIRGKMLITTTKSGDAKTYC 240

Qy 241 VGNVNGERESVLAELTVLERPSFKRPSNLAVTDDSAEFCRAGDPVPTVMKRDG 300

Db 241 VGNVNGERESVLAELTVLERPSFKRPSNLAVTDDSAEFCRAGDPVPTVMKRDG 300

Qy 301 ELKPSYETIRDDHTLKRKYTAGDMGSGYTCVARNYMGKASATLTVQEPHFVVKPRDQ 360

Db 301 ELKPSYETIRDDHTLKRKYTAGDMGSGYTCVARNYMGKASATLTVQEPHFVVKPRDQ 360

Qy 361 VVALGRVTTFQCAZGNPQPAIFWRREGSONLSTFQPPSSRSFSVQGLZLTWNQVR 420

Db 361 VVALGRVTTFQCAZGNPQPAIFWRREGSONLSTFQPPSSRSFSVQGLZLTWNQVR 420

Qy 421 SDVGTYICQLNVAAGSIITKAYLEVTDVADRPPVYRQGNVQVADGTFVLSCVAG 480

Db 421 SDVGTYICQLNVAAGSIITKAYLEVTDVADRPPVYRQGNVQVADGTFVLSCVAG 480

Qy 481 SPVPTLWRKDGVLVSTQDSRIKQLENGVLQIRYAKGLDGTCTASTPSGEATWSAYI 540

Db 481 SPVPTLWRKDGVLVSTQDSRIKQLENGVLQIRYAKGLDGTCTASTPSGEATWSAYI 540

Qy 541 EYQEPGVPPVPPPTDNLIPSAFSKPEVTDVSRNTVLGMPNLSGATPTSTIIEAFS 600

Db 541 EYQEPGVPPVPPPTDNLIPSAFSKPEVTDVSRNTVLGMPNLSGATPTSTIIEAFS 600

Qy 601 HASSGSMQVAVENKVTETSAIKGLKPNATLFLVRAANYGSDPSQSDPVKTQVPLT 660

Db 601 HASSGSMQVAVENKVTETSAIKGLKPNATLFLVRAANYGSDPSQSDPVKTQVPLT 660

Qy 661 SQGVDRKQVRELGNVLLHNPVTLSSSSVEVHWVQDQSYIGTKILYRPSGANS 720

Db 661 SQGVDRKQVRELGNVLLHNPVTLSSSSVEVHWVQDQSYIGTKILYRPSGANS 720

Qy 721 SKWLFEVPTPAKNSVIVPOLRKGVNYELKARFPFQGDSEIKFANTLEERSAPPA 780

Db 721 SKWLFEVPTPAKNSVIVPOLRKGVNYELKARFPFQGDSEIKFANTLEERSAPPA 780

Qy 781 GVTYSKNDGNGTALLVWQPPEDTQNMWQETKWNCLNETIRHINKTVDSGSTSVVPT 840

Db 781 GVTYSKNDGNGTALLVWQPPEDTQNMWQETKWNCLNETIRHINKTVDSGSTSVVPT 840

Qy 841 FLVPGRTISVEAASGAGSVKSEPFQITLDSNGVSPEDQVSLAQISDVKWQAFI 900

Db 841 FLVPGRTISVEAASGAGSVKSEPFQITLDSNGVSPEDQVSLAQISDVKWQAFI 900

Qy 901 AGIGAACWILLWFSIWLIRHRKRNGLSTYAGIRKVPSPFTPTVTVTQGGVSSAG 960

Db 901 AGIGAACWILLWFSIWLIRHRKRNGLSTYAGIRKVPSPFTPTVTVTQGGVSSAG 960

Qy 961 RGLGLLSEPAQWPLADTWNTGNHNDGSIISCTAGNDSNLSLITSRPACDANTYN 1020

Db 961 RGLGLLSEPAQWPLADTWNTGNHNDGSIISCTAGNDSNLSLITSRPACDANTYN 1020

Qy 1021 NQLNDQTNLMPSTVYGVLDLSKNINEMKTFNSNLKDRFPVNSQPTPYATTQLIQ 1080

Db 1021 NQLNDQTNLMPSTVYGVLDLSKNINEMKTFNSNLKDRFPVNSQPTPYATTQLIQ 1080

Qy 1081 SNLSNNNNNGSSGSEKHWKPGQKQVAPQVQINVEQKLNKOTRANDVTPPTITNQ 1140

Db 1081 ANLNNNNNGSSGSEKHWKPGQKQVAPQVQINVEQKLNKOTRANDVILPTIPYNH 1140

Qy 1141 SDYQNTGSSYNSDGRSSTSGSQGHKGAIRPKVPQGGNWDALLPPPAHPPHNSSE 1200

Db 1141 SDYQNTGSSYNSDGRSSTSGSQGHKGAIRPKVPQGGNWDALLPPPAHPPHNSSE 1200

Qy 1201 EYHISVDESTQEMPCVPVPMYIQQDELEEEERDGPPTPVVGAASSPAVYSHQST 1260

Db 1201 EYHISVDESTQEMPCVPVPMYIQQDELEEEERDGPPTPVVGAASSPAVYSHQST 1260

Qy 1261 ATLTPSQDEQLQMLQDCPEETHGMHQPDORRRQVPSVPPPPRISPPHTYGISPLVS 1320

Db 1261 ATLTPSQDEQLQMLQDCPEETHGMHQPDORRRQVPSVPPPPRISPPHTYGISPLVS 1320

Qy 1321 DMDTDAPEEEDAEADMEVAKMTRELLLRGLQETPASSVDGLESSYTGSMINGWSASE 1380

Db 1321 DMDTDAPEEEDAEADMEVAKMTRELLLRGLQETPASSVDGLESSYTGSMINGWSASE 1380

Qy 1381 DNISSGRSSVSSDGSFPTDAFAQAAAAEYAGLVARRQMDAAGRHHFASQCFPR 1440

Db 1381 DNISSGRSSVSSDGSFPTDAFAQAAAAEYAGLVARRQMDAAGRHHFASQCFPR 1440

Qy 1441 TSPVSTDSMSAAVMQKTPAKKLKQPHGLRRETYDOLPPPPPAIKSPTAQSKTQ 1500

Db 1441 TSPVSTDSMSAAVMQKTPAKKLKQPHGLRRETYDOLPPPPPAIKSPVSTAQ 1500

Qy 1501 LEVRPVPVVKLPSMDARTASSDRKSGSYGRVLDGQVVMKNTGDPREARQQMD 1560

Db 1501 LEARPIMPXKLASTEADRASSDRKSGSYGRVLDGQVVMKNTGDPREARQQMD 1560

Qy 1561 KGGNGNAAROLPPAKTHLIQEDILYPCRPPTTSMNPRDSSSSSSSGSGSRQDQA 1620

Db 1561 KAGTITAKHOLPPAKTHLIPEDILYPCRPPTTSMNPRDSSSSSSSGSGSRQDQA 1620

Qy 1621 NVGRHNAEMVVLGGFERGDNNEBLETES 1651

Db 1621 NVGRHNAEMVVLGGFERGDNNEBLETES 1651

RESULT 3

089026

ID 089026 PRELIMINARY; PRT; 1612 AA.

AC 089026;

DT 01-NOV-1998 (TrEMBL: 08, Created)

DT 01-NOV-1998 (TrEMBL: 08, Last sequence update)

DT 01-OCT-2000 (TrEMBL: 15, Last annotation update)

DE DUTTL PROTEIN.

GN ROBOT OR DUTTL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RA Wu M.C., Lowe N., Fordham R., Rabbits P.;

RT "The mouse homologue of human DUTTL/H-robot gene: protein sequence and chromosomal location."

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y17793; CAA76850.1; -

DR HSP: P56276; 1TLX.

DR MGD: MGI:1274781; Robot.

DR INTERPRO: IP000177; -

DR INTERPRO: IP0003006; -

DR PFM: PFM0041; fn3; 3.

DR PFM: PFM0047; fn5; 5.

SQ SEQUENCE 1612 AA; 176406 MW; 5F288C54479684B CRC64;

Query Match 93.1%; Score 8120; DB 11; Length 1612;

Best Local Similarity 95.6%; Pred. No. 0;

Matches 1525; Conservative 33; Mismatches 37; Indels 0; Gaps 0.

Qy 57 TGSRLQBDPFRIVHSDLTIVSGEPATLNCAGRPPTTIEWYKGERVETDKDQPSHRM 116

Db 18 SGRSLQBDPFRIVHSDLTIVSGEPATLNCAGRPPTTIEWYKGERVETDKDQPSHRM 77

Qy 117 SHRMLPSGSLFFLRIVHGKSRPDGSGYTCVARNYLGEAVSHNASLEVALLRDQFNPQSDWM 176

```
Query Match      35.6%; Score 3104; DB 11; Length 1050;  
Best Local Similarity 56.6%; Pred. No. 2.4e-199;  
Matches 613; Conservative 150; Mismatches 252; Indels 68; Gaps 19;
```

```
QY    58 GSRLLRQEDPFRPRIVBHPSDLVSKGEPATLCAKAGRPITPI-----EWTKGGERY 108  
      ||| ||||| | | : ||| ||| | | | | | | | | | | | | | | |  
qh    21 GSRIRORPPDPKKVPKPDSFVTVSKGKPNIPNNWKGGRPPPTIGKVORMVGPDGW---- 75
```

Qy 109 ETDKDDP-SHRMLLPGSGSLFFLRIVHGKSRPDGYYVCAVNLGEAVSHNASLEVAI 167
 Db 76 --TKDDSKVIGCLLPGSGSLFFLRIVHGKSRPDGYYVCAVNLGEAVSHNASLEVAI 133

Qy 168 LRDDPQNPSPDVMVAVGPAVMECPGPRGHPEPTISWKKDGLDDERITIRGGKMLI 227
 Db 134 LRDDPQNPSPDVMVAVGPAVMECPGPRGHPEPTISWKKDGLDDERITIRGGKMLI 193

Qy 228 TYTKRSADGKIVCVGTVNNGERSEVAELIVLERPSFVKRPNLAVTDDSAEFKCEARG 287
 Db 194 SNTKRSADGKIVCVGTVNNGERSEVAELIVLERPSFVKRPNLAVTDDSAEFKCEARG 253

Qy 288 DPVPTVMRKDGDELPKSRYEIRDDHTLKIRKVTAGDMSGYTCAEVMNGKAEASATLTV 347
 Db 254 DPVPTVMRKDGDELPKSRYEIRDDHTLKIRKVTAGDMSGYTCAEVMNGKAEASATLTV 313

Qy 348 QEPHFVYKPRDQVVALGRVTPQCACGNPQPAIFVNRREGSNLLFTYQPOSSSRFSY 407
 Db 314 RAPQFVYKPRDQVVALGRVTPQCACGNPQPAIFVNRREGSNLLFTYQPOSSSRFSY 373

Qy 408 SQGDLITINQVRSDDGYIQQTLNAGSIITKAYLEVTVDIADRPVPIRQGVNQVTA 467
 Db 374 SQGDLITINQVRSDDGYIQQTLNAGSIITKAYLEVTVDIADRPVPIRQGVNQVTA 433

Qy 468 VQDTPVLSVATGSPVPTILWKKDGLVITQDSRIKQNLGVLIYAKLQDGRRTCTIA 527
 Db 434 VQDTPVLSVATGSPVPTILWKKDGLVITQDSRIKQNLGVLIYAKLQDGRRTCTIA 492

Qy 528 STPSGEATWSAYIEVQFQVPPVPRPTDPLNLPSPASPEYDVSNTVITLWQPNLNS 587
 Db 493 STPSGEATWSAYIEVQFQVPPVPRPTDPLNLPSPASPEYDVSNTVITLWQPNLNS 549

Qy 588 GATPTS--IIIFASHSAGSSQTVANWKTETSATGLKPAIPLFLVRAANNGISDPS 646
 Db 550 GATPTS--IIIFASHSAGSSQTVANWKTETSATGLKPAIPLFLVRAANNGISDPS 609

Qy 647 QISDPVKTQVLPSPQGVHDKQVRELGNALVHLNPTVLSSSSIEVHWVDQSQYIQT 706
 Db 610 QISDPVKTQVLPSPQGVHDKQVRELGNALVHLNPTVLSSSSIEVHWVDQSQYIQT 669

Qy 707 YKILYFSGANHGSDMLVFEVTPAKNSVIFDLAKGVNIEIKARFPNFEQGDSEIK 766
 Db 670 YKILYFSGANHGSDMLVFEVTPAKNSVIFDLAKGVNIEIKARFPNFEQGDSEIK 729

Qy 767 PAKTLEAPASQPGVTV-SKNDGNTALYKSPPEPTDNGMVQIKWCLNGNETRYH 825
 Db 730 PAKTLEAPASQPGVTV-SKNDGNTALYKSPPEPTDNGMVQIKWCLNGNETRYH 789

Qy 826 INKTVGDSFTSVIPLPVGIKRYSEVAASGAGSGVSEPPQILQAHGNNVSPSDQVS 885
 Db 790 INKTVGDSFTSVIPLPVGIKRYSEVAASGAGSGVSEPPQILQAHGNNVSPSDQVS 849

Qy 886 LAQIQSDVQKPAFIAGIGACWILMVSINLYHRRKKNGLSTIAGIRKVPSPFTPT 945
 Db 850 LAQIQSDVQKPAFIAGIGACWILMVSINLYHRRKKNGLSTIAGIRKVPSPFTPT 906

Qy 946 TVTYQGV-GEAVSSGGRGLLNSIPAAQWPLADTVPNTGNHNDKSIISCTAGNNGSDS 1004
 Db 897 TVTYQGV-GEAVSSGGRGLLNSIPAAQWPLADTVPNTGNHNDKSIISCTAGNNGSDS 944

Qy 1005 N-LTTSRPAACIANYNQNLNKLMLPESTVYGDVLSNKINEMKTSPNSPMLDGRF 1063
 Db 945 N-LTTSRPAACIANYNQNLNKLMLPESTVYGDVLSNKINEMKTSPNSPMLDGRF 900

Qy 1064 VNFSPQPTPIATTLQSLNKNMNGSDGSEKHWKPLQKQGEVAPQYKINVEQNLN 1123
 Db 991 VNFSPQPTPIATTLQSLNKNMNGSDGSEKHWKPLQKQGEVAPQYKINVEQNLN 1046

Qy 1124 KDY 1126
 Db 1047 NAY 1049

RESULT 5
 Q92214
 ID Q92214 PRELIMINARY; PRT: 1344 AA.
 AC Q92214;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE RIG-I PROTEIN.
 GN RIGI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelaeostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF060570; AAD11628.1; -
 DR HSSP: P56726; 1TLK.
 DR MCD: MGI:134102; Rbgl1.
 DR INTERPRO: IP001777; -
 DR INTERPRO: IP003006; -
 DR Pfam: PF00041; fn3; -
 DR Pfam: PF00047; Ig; 5.
 SQ SEQUENCE 1344 AA; 143439 MW; 890060341C49CEA CRC64;

Query Match 29.9%; Score 2607.5; Db 11; Length 1344;
 Best Local Similarity 38.2%; Pred. No. 6.5e166;
 Matches 594; Conservative 204; Mismatches 456; Indels 299; Gaps 31;

Qy 58 GSRLRQDDPPIVTRHPSDLIVSKGEPATLNKAEGRPTPIEWYKGGSEVTRDADPNS 117
 Db 32 GSRLRQDDPPIVTRHPSDLIVSKGEPATLNKAEGRPTPIEWYKGGSEVTRDADPNS 91

Qy 118 HRMLPGSGSLFFLRIVHGKSRPDGYYVCAVNLGEAVSHNASLEVAIADDERQNP 177
 Db 92 HRMLPGSGSLFFLRIVHGKSRPDGYYVCAVNLGEAVSHNASLEVAIADDERQNP 151

Qy 178 DVMVAVGPAVMECPGPRGHPEPTISWKKDGLDDERITIRGGKMLITTYRSDAG 237
 Db 152 DVMVAVGPAVMECPGPRGHPEPTISWKKDGLDDERITIRGGKMLITTYRSDAG 211

Qy 238 YVGVTVNNGERSEVAELIVLERPSFVKRPNLAVTDDSAEFKCEARGDPVPTVMRK 297
 Db 212 YVGVTVNNGERSEVAELIVLERPSFVKRPNLAVTDDSAEFKCEARGDPVPTVMRK 271

Qy 298 DDELGPKSRYEIRDDHTLKIRKVTAGDMSGYTCAEVMNGKAEASATLTVQEPHFVYK 357
 Db 272 DDELGPKSRYEIRDDHTLKIRKVTAGDMSGYTCAEVMNGKAEASATLTVQEPHFVYK 331

Qy 358 RDQVVALGRVTPQCACGNPQPAIFVNRREGSNLLFTYQPOSSSRFSYQSGDLITIN 417
 Db 332 RDQVVALGRVTPQCACGNPQPAIFVNRREGSNLLFTYQPOSSSRFSYQSGDLITIN 391

Qy 418 VQSDGYIQQTLNAGSIITKAYLEVTVDIADRPVPIRQGVNQVAVDGTFLVSCV 477
 Db 392 VQSDGYIQQTLNAGSIITKAYLEVTVDIADRPVPIRQGVNQVAVDGTFLVSCV 451

Qy 478 ATGSPVPTILWKKDGLVITQDSRIKQNLGVLIYAKLQDGRRTCTIASTPSGEATWS 537
 Db 452 ATGSPVPTILWKKDGLVITQDSRIKQNLGVLIYAKLQDGRRTCTIASTPSGEATWS 511

Qy 538 AYIEVQFQVPPVPRPTDPLNLPSPASPEYDVSNTVITLWQPNLNSGATPSTV 596
 Db 512 AYIEVQFQVPPVPRPTDPLNLPSPASPEYDVSNTVITLWQPNLNSGATPSTV 569

Qy 596 EAFSHSAGSSQTVANWKTETSATGLKPAIPLFLVRAANNGISDPSQISDQVKTQ 656
 Db 570 EAFSHSAGSSQTVANWKTETSATGLKPAIPLFLVRAANNGISDPSQISDQVKTQ 629

Qy 657 VLPSPQPTPIATTLQSLNKNMNGSDGSEKHWKPLQKQGEVAPQYKINVEQNLN 716
 Db 629 VLPSPQPTPIATTLQSLNKNMNGSDGSEKHWKPLQKQGEVAPQYKINVEQNLN 696

RA Bakaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; C
RA Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; C
CX Ephrydiora; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams D.M., Celisier S.E., Holt K.A., Evans C.A., Gocayne J.J.,
RA Amanatides P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.M.,
RA Sutton G.G., Wortman W.R., Vandeell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazee J.G., Champe M., Pfeiffer B.D.,
RA Abel J.H., Doyle C., Baxter E.G., Helt G., Nelson C.K., Miklos G.L.G.,
RA Warr J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew N.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.A.R., Bouck J., Brockstein P., Brottler P.,
RA Burtis K.C., Busan B.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.W., Cavley S., Dahlke C., Davenport T.B., Davies P.,
da Pablos B., Delcher A., Deng Z., Mays A.D., Dev I.I., Dietz S.M.,
RA Dodson K.J., Doup L.E.D., Downes H., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelistola C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris M.L., Harvey D., Belman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegyan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.S., Kodira C.D., Kraft C., Kravitz S., Kulp D., Latz J.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshfiei A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacelci J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen R.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodgate C., Wooley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong Y.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*;"
RL Science 287:2185-2195 (2000).
RX EMBL: AE003458; AAF6887.1. -.
DR HSSP: P56276; ITLK.
DR FLYBASE: FBgn0005631; robo.
DR INTERPRO: IPRO01777; -.
DR INTERPRO: IPRO03006; -.
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; Ig; 5.
DR PRINTS: PR00014; FNTYPEIII.
SQ SEQUENCE 1395 AA; 151758 MW; 25CED70DEBA4F13F0 CRC64;

Query Match 18.3%; Score 1594; DB 5; Length 1395;
Best Local Similarity 30.0%; Pred. No. 5.6e-98;
Matches 419; Conservative 189; Mismatches 482; Indels 308; Gaps 40;

RESULT 6		
Q9W213		Db 56
ID Q9W213 PRELIMINARY; PRT; 1395 AA.		
AC Q9W213;		
DT Q1-MAY-2000 (TrEMBLrel. 13; Created)	Qy 128	
DT Q1-MAY-2000 (TrEMBLrel. 13; Last sequence update)	Db 114	
DT Q1-OCT-2000 (TrEMBLrel. 15; Last annotation update)	Qy 188	
DE ROBO PROTEIN.	Db 173	
GN ROBO.		
OS Drosophila melanogaster (Fruit fly).		

Qy	68	PRIVEHPHSDLIWSKGPALNCKAEGPPTPIIEWGGERVETDKDRPSSHEMLPSGSL	127
Db	56	PRRIEHPDLDVKNKEPALNCKVGKPEPTIENFKDGPEVST---NEKSSRWQFKDGL	113
Qy	128	FFRLIRVHGKSRDPDEGYTVCAVMYLGAEVSHASLEALLRDQFQWSDVMVAWGEPA	187
Db	114	FFRTIRQGGKKQ--DGGEYCVAKMVRQAVSRHASLQJAVLRDQFVEKPTVRKAGETA	172
Qy	188	YMECPFRHGPPEPTISNWKDKSPILD-----KDERITI--RGKMIMTTTRESDAKGV	239
Db	173	LLBCEPPGKPEPTPLIMIKDQVPLDLKAMSGASSRVIYVDGKLLISNVEPTDGNKT	232
Qy	240	CVGTNMGWGRSEVAELTVLERPSFVYKRNSLNAVDDSAEFKCEARGDPVPTVYRKDD	289
Db	233	CIQNMVGTSSSTAKLIVQVQPIYMKPKDQVMLGQTATFHSKVGCGDPQPYVLMKEE	292
Qy	300	GELFASREIE--RDHKLTKRKTVADGKSGYSTCAVMNWKQAZASAILTVQPEPWFKPR	358
Db	293	GNTVPSRIVTFCATGKLEISNTPVDSGTVGNLQVSGNGLQVQISRAASLIYHAPPTNPS	352
Qy	359	DQVVALRGTIPDCKATGPNPALFVHRSQCNLQVSGPQSSPSSPQSGTGLTITVH	418
Db	353	NKCYIVNIVDQLCMASGNPPSPVETKEGVTLMF-----PNSSHGRQIVADGTLTQTV	409

QY	68	PLRVEPESDLSVKSQPAFLNCKAGRPPTITBWKYSGERVETDKDPSRSHMLSPGSL	127
Db	31	PVLIIEHPDILVWVSGSPALNCKAG-PS*AKYITWKQDPVITNKQVMSHRIYDLTSG	69
QY	128	FFLRILYHGKSR- PDGSGVTCVANNIGLGVSHNASLEVALRDFDPNPSDVMVAGEP	186
Db	90	FLFKVSKGNKNGKSDSAGATCYCAENEGEYKSNESGLKMLAEEDPVRPPTVQALGGE	149
QY	187	AWMEQPPFGHPPEPTISWKKDGLPDDKD- ERITING- GKLMIYTRKSDAGKYTCVGN	244
Db	150	AVLECSPPGPGPEPVVSNKDKDELRIDQMPRTILHSGNLIDLPDVSDDSGTQCVANN	209
QY	245	MVGERESVALLTVLERPSVYKPSNLAVTVDSOAFKCEARGDPVTVYKRWKDGELF	304
Db	210	MVGERVSNPRLSVFEKPFKEFGKMTVYCGAALFECRVTPQDGPQITWKRKNEMPV	269
QY	305	SR-YREIDHDLILKRVTDGMSGTCAVENMKGAESATLIVQEPHFVYKPDQOVA	363
Db	270	TRAYIAKNDKGLIERIVPDSGEGVTCYANRAGTLEASAHVQAPSPFQTPADGSP	329
QY	364	LGMVTYQFCATGNQPAIFWRRSGSNLFL- STOPPSSSRESVSQGLDILTNWQRSD	422
Db	330	AGGATATTCILVQSPATFWKSGQGLLFPST- VSDAGRTVSPSTGLILEEDVRQ	387
QY	423	VGTIICQLNWAGSIITKALEYEDVIAD ---- RPPPIVQCPVQVTVADGIFVLSVC	477
Db	388	EGAVICMCHNGAGSSLSKALVTKVATGNTPAKPTITIEHQHQMTLVGSSAILPQ	447
QY	478	ATGSPVPIILMKRGDVLVSTDSRIKLOHSGVLRQIALKGDRGRTCAIATSPSGEATWS	537
Db	448	ASXKPTPIIMWLGRDILVSTDSRIKLOHSGVLRQIALKGDRGRTCAIATSPSGEATWS	507
QY	518	ALIVEQGVGVQPPPTDPTILPSKSPSVTVSNQVQVSNQVNLGAGP-TSYII	596
Db	480	ASLIVEDRTSHAQGVNPDSPNPSSTQPIIVNVTIDEVELHKNAPSTGAGPTGTII	567
QY	597	EAFSBASSGSMQVLAENVKTTESAKGLKPNALFVIRANAGNSIDSPSDISDPVKT-	654
Db	558	OTYSPDLQGVNPIPDVASTYRIKGLKSPSHVMYIRANEGKGTIPSVSVALTSSK	627

Qy 655 ---QDVLPSTSGQGVQVREL-GNAVHLHNPVTLSSSSIEVHTVDQSQSIQTQKIL 710
 Db 628 PAQVALSDSKNNMDMAIAKRLTSQLLKBVKTINSTAVLFWKKRLBELIDSYTK 687
 Qy 711 TR-PSGANGGESDMLVEFVTPAKNSVYIPDLRGVNYEIKARPP-----FNEFGADSEIK 766
 Db 688 WRGPPFTNNQ-----YVWYSPSTENYVSNMPTNYLFFVPIPHSGVHSIGASNSM 743
 Qy 767 FAKTLEAPSPAGPQVTSKNDGNTALLVSQPPEDTQNGWQVETKVCWLGNETRIHI 826
 Db 744 DVILAEAPSPAPPEDVIRML--NLTILRISWAKPADGINGILKAGQIVQAGPNNNR 801
 Qy 827 NKTVDGSTFSPVIFPLPGIRYSEVAZTSGAGSGVSEPOQLDAGNPNVSPEDQVSL 886
 Db 802 NITTNERAASVTLFELVTGMTYKIRVAARSNGGVG-----SHTSEVIMNQDTL 851
 887 AQSISDVVQAFIAGAACWI-----LTMVTSWLYRHKKNGLSTAGIRKVP 939
 852 EKHLAAQKQNSFLYGLINKSHVPPVIVALLIYFVILLIATYWNRSNSD-----GDR 907
 Qy 940 SFTPTPTVTIYORGEVSSGRPLLINSE-PAQPLWADTWPNTGNHNDKISISCTAG 998
 Db 908 SF-----IKINDSVEMASH--NLDVDAQPNQPNMYNTAGIRKMNKNQALYSLTPN 959
 Qy 999 WGNDSNLTITS-----PRADCIANTNNDLNKOTNLMLPESTYGVDSLNKINEXKTEN 1054
 Db 960 AQDFQKNCDTSGTHMRPGEHRYHAQITGPGGNM--SFTYG-----MYRND----- 1006
 Qy 1055 SPNLKAGDFRPNVSGQPTPIATTLQSLNNMNGSGDSGEKHWKPLQGVQVAPVQY 1114
 Db 1007 -----DPSPTATTILVLSN-----QQ-----PAML 1026
 Qy 1115 NIVQZNNKLKDRANDTVPPIYVNSQVDNTGGSVNSDRSS-----TSSG 1162
 Db 1027 NM--DKMLRAPMNPVPPPPPP--ARYADTAGRRSSRASDGRGLTNGLHHTSSG 1081
 Qy 1163 Q-----GHKK-----GAPKPKYGGGMMADLLPPPPHPP 1195
 Db 1082 QRSQSPHTDVSIVQLASSDGTGSSKERTGRTT--PNYLM--DPPIPPYSPNPP 1134

RESULT 9
 Q9V266
 ID Q9V266 PRELIMINARY; PRT: 859 AA.
 AC Q9V266:
 01-MAY-2000 (TrEMBLrel. 13, Created)
 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 GN C55423 PROTEIN (FRAGMENT).
 SN C55423.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM W.A.
 RC STRAIN=BEKLEI;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Annatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-R.C., Blasey R.G., Champs M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 April J.P., Agabeyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Ben C.P., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-S., Ibegun C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merlugo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.N., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000)
 DR EMBL: AB003586; AAF51388.1; .
 DR HSP: P56276; 1TLK.
 DR FLRBASE: FBgn0031328; C55423.
 DR INTERPRO: IPR001177; .
 DR INTERPRO: IPR003006; .
 DR PFM: PF00041; fn3; 3.
 DR PFM: PF00047; ig; 5.
 FT NON_TER 1 1
 SQ SEQUENCE 859 AA; 93916 MW; 5CFD69D984101BF8 CRC64;

Query Match 15.0%; Score 1307.5; DB 5; Length 859;
 Best Local Similarity 39.9%; Pred. No. 4.2e-79;
 Matches 300; Conservative 152; Mismatches 345; Indels 87; Gaps 22;

Qy 68 PRIVEPMSOLINSKGEPAATLKAAGRPPTIENWYGGKRVETDKDPRSHMLPGSL 127
 Db 1 PRIVEPMSOLINSKGEPAATLKAAGRPPTIENWYGGKRVETDKDPRSHMLPGSL 57
 Qy 128 FFLATVHGKSRDGGVTCVARNILGSAVNSASLAEIADRDQFNQSPVWVAGPEA 187
 Db 58 FFLATVHGKSRDGGVTCVARNILGSAVNSASLAEIADRDQFNQSPVWVAGPEA 90
 Qy 188 WMEQPPRSHPEPTISWKKDGSPLD-----DKERITIRGKLIMITYKSDAGKTVCGTN 244
 Db 91 LLECAAPRSHPEPTISWKKDGSPLD-----DKERITIRGKLIMITYKSDAGKTVCGTN 149
 Qy 245 WMEQPPRSHPEPTISWKKDGSPLD-----DKERITIRGKLIMITYKSDAGKTVCGTN 302
 Db 150 WMEQPPRSHPEPTISWKKDGSPLD-----DKERITIRGKLIMITYKSDAGKTVCGTN 209
 Qy 303 PKSRITRTHDRLKINKVITAGDSMTYCAENWYKRAESATLTVQEPHVPKPDQV 362
 Db 210 PKSRITRTHDRLKINKVITAGDSMTYCAENWYKRAESATLTVQEPHVPKPDQV 269
 Qy 363 ALGRITVQCEATGNQPAIFWRRRSGNQLFSYQPOSSSRFSVSGD-----LITIN 418
 Db 270 ALGRITVQCEATGNQPAIFWRRRSGNQLFSYQPOSSSRFSVSGD-----LITIN 327
 Qy 419 QRSVDSTT-QDTLWAGSILTKALEYEDVIADRPVPIRQGVQVQVAVDGTFLVSC 477
 Db 328 QRSVDSTT-QDTLWAGSILTKALEYEDVIADRPVPIRQGVQVQVAVDGTFLVSC 386
 Qy 478 ATGSPVPTLNRKLDGVSTQSRQLENGVLQIR-YAKLGDGRITCIASPTSGEATW 536
 Db 387 ATGSPVPTLNRKLDGVSTQSRQLENGVLQIR-YAKLGDGRITCIASPTSGEATW 445
 Qy 537 SAYIEVQGVQVPPPTPDNL-----IPSAKSPKPTVDTSRNTVLSNQPLN 586

Db 446 SGFLRE-----LPTNPNKIFYRAPEQTKPSAQOQPKLLNATASALITWPTSDK 496
 Qy 587 SGATP-TSYIEAFSHAGSSGWTVAENKVTETSAIKGLKNAYILFLVRAANAYISDP 645
 Db 497 AGASSPFGTSGVMYCTQSRWTPIASLSEPIYTESIQGAAMFIVRAENSLGSPFP 556
 Qy 646 SQISDPKVTQDPLTSGQVDRHQVREL-----GNVLHLNHLNPTLVSSSIEVHT 696
 Db 557 SPISEPIAGKLVGRDSESTGTSQLLSVETLQANDVLELANASDSTARLSWD 616
 Qy 697 VDOOSQIYQIKLYRPSGANNESDNLVFEVTPAKNSVPIPLDRNGVMEIKARPFEN 756
 Db 617 ID-SQYITGGFLYARE--LHSEYKMWLLNKGQSLCTVGLAXASTYEFVLPYTK 673
 Qy 757 EFGQADSEIKFAKLEAPSPQGVTSKNDGNTALLVSQWPP-PEDTQNGVMYQYK 815
 Db 674 SIVKPGNSNRHMTLDVPEAPPYGMBAIQ--PWRYSVWLKMLPPQNRKNGILLSYW 731
 Qy 816 WCLG-----NTRYHINKYVDSSTFSVIFPLPGIKRYSVEASTAGSGVKSEPFQIL 871
 Db 732 VVKGLDWHVTRIFKNMTIDAAPILLANTIGVTYIVAAVATVGVQPFSPKAVLRI 791
 Qy 872 DABGNPV-SPEDQVSLAQGIS-DVVKQPAFAGIAGACKIILMY 913
 Db 792 DARTQSLDGVYTRIPSRDADDFLQTVFLVLSIAITVFL 835

RESULT 10

Q9VQ10

ID Q9VQ10 PRELIMINARY; PRT; 823 AA.

AC Q9VQ10;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CCS481 PROTEIN.

GN CCS481

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Annatides P.G., Scherer S.E., Li P.W., Hoshins R.A., Galle R.P.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.J.,
 Brandon R.C., Rogers T.H.C., Blazek R.G., Champs M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Frankoach C., Baldwin D.,
 Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 Burtis K.C., Susan D.A., Butler R., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov C.A., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegwan K.,
 Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Krevitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Wellenbach J.,
 Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.T., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: A2003586; AAF51373.1; .
 DR HSP: P56276; LTKL.
 DR FLYBASE: Fg0031341; CG5481.
 DR INTERPRO: IPR001412; .
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFM: PFM0041; fn3; 1.
 DR PFM: PFM0047; 19; 5.
 DR PRINTS: PR00014; FMTYPIII.
 DR PROSITE: PS00176; AA_TRNA_LIGASE_I; UNKNOWN_1.
 SQ SEQUENCE 823 AA; 89715 MW; 36FC0B91F362F19 CRC64;

Query Match 13.94; Score 1212; DB 5; Length 823;
 Best Local Similarity 33.44; Pred. No. 9.8e-73;
 Matches 279; Conservative 127; Mismatches 273; Indels 156; Gaps 21;

Qy 76 DLIVSGEAPLCKAKGRPTTIEWTKGRVERTDKDDPSRHLPLGSLFFLIVNG 135
 Db 2 DTTVPANDPFTQCAENGNPTTQWKGRELKTDGTG--SHRLLPAGSLFLAVINS 58
 Qy 136 KRSRDPGVTVCAVNIAGVASHNASLVAILLRDFQNPDSVMVAGPEYAWQPPR 195
 Db 59 R-ESDAGTWTCANKEFGVARSNATLQVAPLDRFLPEANITVAGQVLAEMAGAP 117
 Qy 196 GHEPPEISKNKGSPLD--DDEXRITINGKMLITTYRSACKYVCGVNMWGERSE 252
 Db 118 GSPFPEISKNKGQTLNKGKRII-VGGNLQIAEQSGDGTQCVVNVNGTRESA 176
 Qy 253 VAEITVLEPSPVFRKSNLAIVTDSAEFKCEARGDPVPTVNRK--DQGLPKSRYEIR 310
 Db 177 TAFILVNVPRPILRGPQNTAVGSSVPTQCRIGDPLDPLVLRATASGCMR-PRVVL 235
 Qy 311 DORTLAKIRVADMGSTVCAENMYKAEASALIVQEPHFFVRKQDVVALGRTYTF 370
 Db 236 EDRSLADLVTDLMGDEGTCEADNAVGGITATGLVTHVAPPEFVRKMGVLGIDVLF 295
 Qy 371 QCATQNPQPAIFWRRGQSNLFYQPOSSSRFSVSTGQD--LITINQVRSQVY 426
 Db 296 EQCANGRPPTLVKSVGNSGLL--PYGRGMEVTLTPEGSVLIRAFRDGSKV 352
 Qy 427 I-QYLVAGSIITKALEVTDVIADRPVPIRGVQNVTVADGTVFLSVAGVSPYT 485
 Db 353 VTCNALNAGVSSRTVSVY-DIQEPLPTEIGQVNTLPKVSIVLPCRTGLPVPQ 411
 Qy 486 ILMRKDGVLNVTQDSRIQLEN-GVLQIR-YAKLGDGTGYCTIAPSPGEATSWAIVQ 543
 Db 412 VSMVLGIDIPVQDEERNLSDAGALTISDQREDGLTICVANSRNGKSSNGSLRLD 471
 Qy 544 EFGVQVPPRPTDNL-----IPSAPSKPEVTVSRNTVLSN-QPNLNSGATPT 592
 Db 472 -----TPTNPNKIFYRAPELSTYVGPCKPQVKEGNSVLSWTSNKGSSGLV 522
 Qy 593 SYLIEAFSHAGSSGWTVAENKVTETSAIKGLKNAYILFLVRAANAYISDPISDP 652
 Db 523 GYIYEMFGKNETDGAWAGTVQNTPTTQGLPGVNTPLFRAENSGLSLSPSEPI 582
 Qy 653 KTDQ-----LPT-----SQGVDRQVQRE-LGNVLEHLNPTLV 686
 Db 583 TVGTVSSSEHNSPTFMPSLHIFDLSFPQRTFNSGLDSEARSLISGQDVELSNASV 642
 Qy 687 SSSSIEVHTVDOQ-----SQYIYQIKLYR----- 712

0... 264 TGGTGTGCGATACAGCCTATTGTTGGCAAACTA CAGGAGGAGGTCGACGATCTATGAGCC 429

QY	70	IIVHSFSLVSKGPEATLACKACRGPTTPIIWKYGKRVETIDKDPPSHMLLSGSLEF	129
Db	67	LVE-PVDTLVRSGSSWILMCASIEPSPNIEMKKDGTFLALEDSD-----RRQLLGPSLEI	122
QY	130	LAIRVAGRSKPDGEVTCVCVAR-NITLGEAYSHASALELIRDORFNPSDMVAWEPAV	189
		: : : : : : : : : : : : : : : : : : :	
	123	SNWBSHKNRPGEGQCFAVDMLGTVISRTAKTIVGL-PRTPSQPEPSSVYGCNSAI	181
QY	189	MEOCPPRGHPEPITSWMKDGSLDDKERITRGKLIMITYTSIKDSKAGTCVGMNMGSE	248
Db	182	LNC-EVNADVPFYRWEDNRGLLLDRDIKVLSPSTLVINATGEDGGLYLCITYESGPDP	240
QY	249	RESEVAELTIERPSS-----FVKRPSNLAVTDSDASEFCBEARGPVTPIVMRKDDGEL	302
		: : : : : : : : : : : : : : : : : : :	
	241	KFSDEAEIKVLQDPPEITVDLPFARPMSSMKVTQSVALPCVYSGVAPVMMNMKEVLT	300
		: : : : : : : : : : : : : : : : : : :	
	359	-----PKSRTEIKDDHTLIKRVTAGDMGSTVCARMNWGAASALITVLEGPHPVYKRPD	
		: : : : : : : : : : : : : : : : : : :	
	360	DTESGRVLVLAGGLEISDVYEDDAGTYFCIADNGNKTYEAOQLVLPVWPGLAQPAN	360
QY	360	QVVALGRTVTFQCEATGNQPAPFWRREGSONLFIYSQQPSSSFRSVQTGDLETITNQ	419
		: : : : : : : : : : : : : : : : : : :	
	412	IYAHESHVIDFEVCFTGKPTPIVWYKWDVI-----PSDNFKIVKHNVQLGVLI	412
QY	420	RSDVGTYICQFLNVAIGSIITRAYLEV--DVIAEDPPPVRQPGPQVAVDGLFTVSCV	477
Db	413	KSDGEPYQCIARMVDNQAQOILLHEHDVAITPLP-----TSLISTADTHLAP	463
Qv	478	ATGSPVPIILMRKGDLVSTOSRIKOLENGVLOIRTAKLDGTGYCTIASPSTGW-A	536

Db 464 ATGTCPSAPADVVASLVT-----RFLK-----TWRTPSADPHONLWY 504
 QY 537 SATIYQVEGVQVPFPPRTDPMILPSASKPEVDYSKMTLWSQWNLMSGATPTIYI 596
 Db 505 SVYITE--GYDRKRENT-----SQGEMQVT----- 530
 QY 597 EAFSHSSGSSQWTVAENKWTESAIGLKNAIYLFLVRAANAYIGDSQISDVPKVTQ 656
 Db 531 -----TQNLMPATVIYFKVMQNKKG--SGESSAPLRVETQP 565
 QY 657 VLPTSGQVGHQVRELGNALHLNPLTVLSSSSIVHW--TVDOQSQIYQCKILYLRPS 715
 Db 566 EY-----QLPGCAPINRIATSP-----SITVTEPLSGSEIQTMLKTYMKG 614
 QY 716 ANNGESDMLVEVPTPAKMSVYIPDLRAGWYIEIKARPFNEFGGADSEIKFARTELEAP 775
 Db 612 YDK-EQDVI-----SSSITINGLKAYTISFPAVNAIKHKGPGVSTQDVAVRLSDVP 664
 QY 776 SAPPGQVTISKNDGWTALVWSQPPEDTQCNQWVQETKWCNLGNETRYHIMKT--VDGST 834
 Db 665 SAAPQNLSEVR--NKSIVHWQPPSTTNGQGITQYKIRTKRAKSDVETLWGTQ 722
 QY 835 FSVYIPFLVGIIRYSVEVAASTGAGSCKSE-----PQIFQLDAGHWPN 878
 Db 723 LSQLEIGLRDGTETNFRVAALVNGSTGATDWSABTPESDLETWYVP--VSSHLVRLP 781
 QY 879 -----SPEDQVSAQDISVQKPAFIGAIGACWIILMYSWLYTRHKKNGKLT 929
 Db 782 VTSIVSYMPTFEN-----NLVGRGTAIGYIGSPHAQIKVD-----YKGR----- 823
 QY 930 STYAGIRKY-PSFTTPTV-YTQRGSAV-----SSGGRP-----GLNISEP-AAQWP 975
 Db 824 ---TYTLENLDSSHTVILAKANNVGBGILPYLESAIVRPHDITSEVDFVINAPITVPD 881
 QY 976 LADTWNTIG-----NNNDCSISQCTAGNCSNLSLITVSRPACDIANTNWNLDNQTNL 1030
 Db 882 PIMPMPPGVQASILSHDIRTW-----ADNSLPKRRQKTO--SRITYV--RWKTN-- 929
 QY 1031 MLPESTVIGDVLNKNEMKTPNSP-----LKQGRFVNPSCQPTPYATQLQSN 1082
 Db 930 -IPANKTKNAN-ATTLSYLVGLKPLNTLIESVMTKGRSSWTSMAGATFELVPTP 987
 QY 1083 LSNNNNGSGDSGEK-----HNKPLGQKQEV-----APVYNIVEQ--NKLN 1123
 Db 988 PKPDVTVYSKSGKPTLIVNWQVSPRANGKITGYIYLSSTDVNAELHDWIEVPVGNRLK 1047
 QY 1124 KOTRNDTPVPTIPYNQSTQNTG-----SYNSDR--GSTSGSGKHGK 1168
 Db 1048 --HQIQLDTPITFYKIQARNSKGMPSAVQFRTPKDSSDKMDQLMSAG--KG 1103
 QY 1169 ARTPKPVKGGGNWADLLPPPAHPPHSHNSE-----EYNT----- 1204
 Db 1104 SRLPOL-----GSDYKPMGSGNSPHGSPPTPLDSNMLMLVIVSGVITVVVVVYAV 1156
 QY 1205 -----SVDESDQEMPC-PVPVPMYQLQDELEEDERGPT--PVR 1244
 Db 1157 PCTRRTTSHQKKRAAKCSVNGSHYKGNCKYDQKPLDMIHRELKLPDKSDQPNPV-- 1215
 QY 1245 GAASSPAAVSYSHQSTATILPSPQEEQLPMQLQDCEPTEGHMQ--HQPO----- 1290
 Db 1216 -----MTDTPIPNSQDPT--VDNMSQSHHQRNRSTGRHSSEDSSTLAC 1260
 QY 1291 -RRQQVSPFP--PPRPSPHYTYIGSLPVDMDTDAPEEEDAEEMAVKAMQTRR 1345
 Db 1261 RRGMRKPMXMPDSQPPQPVISAH-----PTHS--LDWPHHHSFSSL----- 1301
 QY 1346 LLRLGELQTPA-----SSWGDESSVIGSMINGWSASEEDNISGRSSVSSDGSFF 1398
 Db 1302 -----ASPRASHLYHPSPPWTGTSN--SLSDRANSTESVMTPTSTDMPASSGTC 1352
 QY 1399 TDAPFAQAFAAAAYAGLVKARRQMDQAGRRHSAQCPRPTSPVSTDSNMAVAQMKT 1458

Page 13

```

      :      :  ::  || :||:| | | | :||::: | : || |
Db 422 Q:----PARILKID--NQITMAVEGSTATLLCKAGAPVPVQWLDDEGTTVLQDERFF 475

      :      :  ::  || :||:| | | | :||::: | : || |
Qy 505 LENGVLQIRAKLGDITGRITCIATPSGEATSWISIEVQEGFVQPPRP----- 555

      :      :  ::  || :||:| | | | :||::: | : || |
Db 476 YAGTSLISDPLQDAGTGRFCQAGANDQNNVILANLQVKEATQIQGPRSAIEKKGARVT 535

      :      :  ::  || :||:| | | | :||::: | : || |
Qy 556 -----DPLN----- 559
      ||:|

```

```

QY 560 -----IPASRKSPEVTD-----VSRNVTLSWQPNLNGSATPYSIIVL-AFSH 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 LDEVESRAQLVLWSPGPHLEDRHLKQSQVLHWSPADHSHPKTEKLEDFEDKE 655
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 602 AGSSGSMQIVANWKTETSIAIKGLAPNAIVLFLVRAANAGISDPSISDPVKTKOVLPTS 661
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 MAPEKTSGLGPNQWQSTLKLSPYTHIFPRTAIKNTGGEPSVSTVTFVPEAPEK 715
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 662 QGVHDQVQRELGNALVHLNRPVTLSSSSIEVHWVDQSQSYIGTKYILRPSGANHGDS 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 716 NPVDVRGBGNTNIMTY-----TWPLRN-MDNNAQIQ-TRVWQPGQK--OE 760
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 722 DNLVFEVHTPAKNSVYTPDLRGVNYEIKARPNFEFGADSEIKFAKLEAPASPP-- 779
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 761 TWREQVTSDP-----FLVSVSTISFPYBEIKQVANNQKGGEPQVITGSGEDYPOSWEL 817
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 780 QGVTVSKMDNGTIALVSNQPPPTDQGMQVEYKV--WCLGNETRY--HINKT--VD 831
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 818 EDIIIF--NBSVTLVRRPDLAQVGHKLGYNTVTKWGSQRKSHKRIHSHVPP 873
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 832 GSTFSVIFLVPGLIRYSVEVAATGACSGVKSE-----POFIQDANG----- 875
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 ANTTSAILSGLRYSYHVEVQAFNGRLGPAESWTFSTPBGVGPPEALRLKQSDTSL 933
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 876 -----NPVSPEDQVLAQQISDVVQKPAFAGIACRWILMF 914
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 934 LLHWPQVISHGVNGLQGLVSLVPGVGRSKLPENSD----- 971
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 972 -----PELRTHNLTNLNPDQ----YRFQLQATTQCGPGEAIVREGGTMALFGKPDFGN 102

```

907 |T|G|A|C|A|T|C|A|T|P|A|L|N|K|N|L|S|C|C|A|T|G|G|G|G|G|G|G|T|T|S|P|A|C|I|A|T|H|N|   102
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1022 |S|A|T|A|G|E|T|S|V|V|W|F|R|K|------Q|N|C|F|R|H|L|K|A|L|P|G|K|G|V|D|P|O|P|Q|V|T|S|W|G|S|T 107
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 1022 -----Q|L|N|K|Q|N|L|M|P|E|S|T|V|G|D|L|S|-----R|I|N|E|M|K|T|F|N|S|-----105
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1078 Q|N|N|Q|P|D|T|K|T|H|L|I|K|E|V|L|L|H|D|V|A|T|N|G|T|G|P|V|T|S|T|G|S|A|S|E|G|W|I|A|F|Y|S|A|I|L|L|L 113
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 1056 -----P|N|K|D|G|F|-----V|N|P|S|Q|P|T|P|A|T|L|Q|S|L|N|N|M|N|G|S|D|G|S|K|- 109
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1138 I|L|L|I|C|I|K|R|S|G|G|K|T|S|V|K|D|K|E|T|Q|V|D|S|E|A|R|P|M|K|D|E|T|G|R|S|L|E|D|N|E|K|A|F|G|S|S|Q|P|L 119
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Qy 1098 ---H|W|K|P|Q|Q|K|Q|E|V|-----A|P|V|Q|---I|V|E|Q|N|L|N|K|Y|D|R|--N|D|V|P|T|P|I|N 1139
      |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1198 N|G|I|K|P|L|D|S|L|A|D|T|G|S|V|D|Q|P|N|E|D|G|S|F|I|Q|Y|S|G|K|K|E|A|A|G|G|D|S|G|A|T|S|I|N 1253
      |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 15
Q90610
ID Q90610 PRELIMINARY; PRI; 1443 AA.
AC Q90610;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI TaxId=9031.
```

```

Db      606 STQWVVRVLSTVSDAPQNLLEAR- :XKSLMLNGWPPAGCHSQITGKIRYKRV 663
Qy      822 TRTHINKTGVSDSESWIPVLPGVIRKSEVANSACGAGS-----VKSPQIQLD----- 872
Db      664 RKSVDSESVGGTQLQLLEGLERGTYTFNFIATMVTNGGTATDWVSATFESDLDSEKV 723
Qy      873 -----AGNVP-----SPEDQVSLAQSDSDVVKQAFIAGIACWIIWVSIW 917
Db      724 PEVPSSLHLVRPLVTSIVSWMTPEMQ-----NIVKGTAGTIGGISPHAQITKVD----- 773
Qy      918 LYRHKRKNKGLSTYAGIRKY-PSFTTPTPV-TYQRGSEAV-----SSGRGP-----GL 964
Db      774 -YKQR-----YTTIENLDPSHHYITLKAFNNGGGLPESLSEAVITRPHSDTSEVDL 823
Qy      965 LNISEP-AAQPLADPTWMTG-----NNNDCSIQCTAGNNGSNDLNTTSPACDIAN 1018
Db      824 FVINAAPTVPVDPSPMPMPGVQASILSDTIRITW-----ADNLSKPKMKTID--AR 874
Qy      1019 YNNQLNQTMLMLPESTVYGDVDLS-----KINEMTFNSPLNDKGRFVPSGPQ 1070
Db      875 YTVV--RWKTK--IPANTKIYKANATTLSVLTGPKLPLTFESVMTKGR--SSTWSM 928
Qy      1071 TPYATT-QLIQNSLNMYNGSDEKSEK-----HWKPLQKQJQEVAP-VQYNVEQNKNLK 1124
Db      928 TAGGTITFLVPTSPKKDVTYSKSGKRTIIVWQGPSPBANGKITGTIITTYSTVDNAEIH 986
Qy      1125 DFRANDTVPPTIPYN-QSYDQGTGGSYNSDRSGSTSGSCHKKGARTPKVKGKMGWMA 1183
Db      989 DWVIEPVVGNHLTBOQLTLDTPTTYFKIQARNSKGMGPMSEAVQPTPKAES-----S 1042
Qy      1184 DLLPPE-----PARPPPHNSSEEVNIYDSESYDQEMPC----- 1217
Db      1043 DRMPNDQAQSGSAGKSRPVDGPDYKPLPGLSGNSPHGSPSTPLDSNMLLIIVSVGVITI 1102
Qy      1218 -----VPPARVYLQDLEEEEDENGCTPPVGRASPSAAVSTYSHQSTALMTPSQDEQLMQ 1275
Db      1103 VIVIVYAVCFTRRTTSQWKKRAACKSVNG-----SHYKNGKSDVWPPDLMI----- 1150
Qy      1276 QDCPECTGHWQHQPDQRARQVSPFPPPPRPFI-----SPHYTGITSGPLVSMDDIAPF----- 1328
Db      1151 -----HHERLELKPDKSPDPNMTDPTIPRANSDPTIPVDNMSDMSIHQARNS 1199
Qy      1329 -----EEEDAEIMEVAKQTRRLLLGLEQTPASSY-----GOLES-- 1364
Db      1200 YRGSEEDSMSITLAGRGNKPKMMMPFDSQPPQVISAHPISHLNDPHHFFSGSLASPT 1259
Qy      1365 -SVTSGMNGW--GSAEEDNISGSRVSSSDSGSFFTDADFQAIAAAAA-----YAGLKV 1418
Db      1260 RSYLHQQVSPMPVGTSMHSDRANSTSEVNTPSSDTIPASSSQPCADHQDPSGSAYLI 1319
Qy      1419 ARRQMQDAAGRHSPHACQCRPTSPVSTDSNMASVAQMTKIPAKKLKHQPHGLRRTTID 1478
Db      1320 GSAQDEIDA-----QSLTAPHRPSBLKSPFAVPAVPAAGSAYDP 1359
Qy      1479 DLLPPEPV-----PPPAIKSPAQSKTQLEVR--PVVYPLPSMDARTDRSSDRKGS 1527
Db      1360 TLPSTPLTLTQAQSPHPVSHVK--TASITGLTRGTRPMPMPVVPSPADVQ-ETTRMLEDES 1416
Qy      1528 SYKRGVE 1534
Db      1417 SEPEDEL 1423

```

Search completed: January 22, 2001, 12:54:03
Job time: 2044 sec

QY	57	TGSRLRQDEPFR-RIHVESPSLIVSGKCAPATLCKHAGRPPTTIEWKGGSRVETDKDDP	115
Db	9	TGSGVVR--TPTTPFTFLEPMDDLVSYGASVIMNCSSCTETPPALNWKDQDGLLWVSD-	65
QY	116	RSRHLMLSGSLFPLFIVGKRSKDPBGVYCVAR-NYLGEASHASLEVALLRDFRQ	174
Db	66	---RQQLDPDGLSLINVSVKHNRKPDGVEYQCVATVESLGSIVTAKTLVAGL-PPFTS	122
QY	175	NPDSWVAVGPAVMEQCQPPHGHPETISWKKDSPDLKDERITINGKMLITTTKRS	234
Db	123	QBELSVSKGSAIALNCE-VNMDLAPFVMEQDQPSLSDRDVFLKPSGALLIGNATD	181
QY	235	AGKYVCVGNMVGRESVEAZLTVLEPSS-----FVKRPSNLAVTDDSAEFCEARG	298
Db	182	GGTFRCVIESGGKYPSAEAKLILPOEPESQVLFVQPSGLITVTGQANFPCVAGG	241
QY	289	PVPTVMRKDDGEL-----PKSYEKIDRHTLKIKRIVTAGDMSGYTCVAENWGKAESATL	345
Db	242	PTPIPVMTNKEELITEDSERFALAGSGSLISDVTEDVGTITCIADNENMTIAQAE	301
QY	346	TWQEPFHPVFKPQDVVALGRVTTPQCAEIGNPQPAIFWARGESQNLSTQPSQSSR	405
Db	302	AVQVPEFLKAPAMITAHESMDIVFCEVTKRPTPTVYKWNQDVIPSDT-----F	353
QY	406	SVSQGDGLITINQVSDVGYICQLNVAGSITIKATLEVIDVIADRP--FPVIRQGPVN	463
Db	354	KIKENKVLQGLIKVSGDEFTYCIAEDNVNGAQAQALLDLDVAIPLPTSLTSATN	413
QY	464	QTVAVDQVSLKSCVATGSPVPTILNRKDQVLVSTQDSRIQLENGVLIRIAKGLDTGRY	523
Db	414	DLHA-----PATGLPLFAPRDVVALVST-----HFLR-----	443
QY	524	TCIASTPGEATWSATIEYQBFVGVPVQPPPTDNPILPSAPSKPEVTVDSRNTVLSWP	583
Db	444	-----TWR-----TVSDQ- GDMLTSLIPTKE- GINRSVERNSP	479
QY	584	NLNSGATPTSYIIAESHASGSSQVVAENKTESA-IGKLAPMAILFVLRAANATGI	642
Db	480	G-----ETWQKQMLQAPETVTFVFNKQKQH	507
QY	643	SDPSQISDPKTVQDVLPTSQGVNDKQVRELGNALVPLVSSSLVSEV-VDQQS	701
Db	508	GESSA--PLK-----VATQPEV--QLPPAPNIRIAYAGSP-----SVTVVTPPLSGN	552
QY	702	QITQKYLKLPKSGANGHSGSDLPVETPAKNSVVIDPLKAGVNTIEKARFPNTGSA	761
Db	553	GEIQKWLKLYMEXGQD-SQDQVDV-----AGLSYTIQGLKKTETSPRVNAVHNKGEP	605
QY	762	DSIEIFKAPLIEASAPPGOVTVSKNGKNTALYSWOPPEPDTONGWOEYKYNWGLCE	821

GenCore version 4.5
Copyright (c), 1993 - 2000 CompuGen Ltd.

ON protein - protein search, using sw model

Run on: January 22, 2001, 12:19:37 ; Search time 233.01 Seconds
(without alignments)
63.689 Million cell updates/sec

Title: US-09-540-245A-19
Perfett score: 2280
Sequence: 1 QIQAQGRVITPPCETKGNPO.....TSAALSQSQRPRPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 268485 seqs, 34139795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseg_36.*
1: /SIDSL/gcgdata/geneseg/geneseg/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseg/geneseg/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseg/geneseg/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseg/geneseg/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseg/geneseg/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseg/geneseg/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseg/geneseg/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseg/geneseg/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseg/geneseg/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseg/geneseg/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseg/geneseg/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseg/geneseg/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseg/geneseg/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseg/geneseg/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseg/geneseg/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseg/geneseg/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseg/geneseg/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseg/geneseg/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseg/geneseg/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseg/geneseg/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseg/geneseg/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2276	99.8	434	20 Y13567	Human Robo 2 poly
2	2276	99.8	434	20 Y08405	Human partial ROBO
3	914	40.1	1649	20 Y08404	Human ROBO1 protel
4	913	40.0	1651	20 Y13566	Human Robo 1 poly
5	877.5	38.5	753	20 W83927	Human T85 protein.
6	561.5	24.6	1297	20 Y13565	C. elegans Robo po
7	561.5	24.6	1297	20 Y08403	C. elegans ROBO pr
8	545	23.9	1395	20 Y13563	Drosophila Robo 1
9	545	23.9	1395	20 Y08401	Drosophila sp. ROB
10	473.5	20.8	1380	20 Y08402	Drosophila sp. ROB
11	473.5	20.8	1381	20 Y13564	Drosophila Robo 2
12	324.5	14.2	1911	16 R71726	Human PTP-OB. Hom

13	324.5	14.2	1911	18 W27225	Human protein tyro
14	324.5	14.2	1911	20 W94027	Human protein tyro
15	323.5	14.2	1501	16 R72858	Rat receptor type-
16	308.5	13.5	1897	21 Y81785	Human protein tyro
17	308.5	13.5	1897	21 Y56100	LAR tyrosine phosph
18	305.5	13.4	761	17 R92255	Neural cell adhesi
19	285.5	12.5	1192	19 W57900	Protein of clone C
20	285.5	12.5	1299	21 Y40439	Human Nr-CAM prote
21	285.5	12.5	1496	20 W81030	Melanoma associate
22	285.5	12.5	1496	21 Y70469	Human p53 target a
23	276.5	12.1	582	17 R92256	Neural cell adhesi
24	268	11.8	1257	20 W74152	Human L1 cell adhe
25	267.5	11.7	1028	19 W29667	Homo sapiens DL185
26	264.5	11.6	1242	19 W52287	Rattus norvegicus
27	262.5	11.5	1304	19 W59994	Human neural cell
28	254	11.1	868	17 R92717	Mouse muscle-local
29	253	11.1	869	18 W26611	Human muscle-speci
30	253	11.1	869	18 W26506	Human Dmk receptor
31	250.5	11.0	867	19 W62583	Mouse receptor tyr
32	250.5	11.0	871	17 R84087	Nsk2 receptor. Mu
33	250.5	11.0	871	19 W62568	Mouse receptor tyr
34	250.5	11.0	881	17 R84091	Nsk2 receptor with
35	250.5	11.0	881	19 W62572	Mouse Nsk2 (altern
36	250	11.0	1091	18 W41641	Sequence used in d
37	250	11.0	1091	20 Y08010	Mouse LIG-1 protel
38	249	10.9	860	17 R92716	Mouse muscle-local
39	249	10.9	1447	16 R68553	Deleted in colorec
40	249	10.9	1447	20 Y33498	Human DCC protel.
41	249	10.9	1728	12 R13144	Deleted in Colorec
42	247.5	10.9	1018	15 R63759	Human contactin (E
43	247.5	10.9	1018	17 R87028	Human contactin.
44	245.5	10.8	863	17 R84088	Nsk2 receptor with
45	244.5	10.7	863	19 W62569	Alternatively spli

ALIGNMENTS

RESULT 1

Y13567
ID Y13567 standard; Protein; 434 AA.
XX
AC Y13567;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 285
FT /label= unknown
FT /note= "encoded by GTW"
FT Misc-difference 396
FT /label= unknown
FT /note= "encoded by NTT"
XX
PN W09925833-AL.
XX
PD 27-MAY-1999.
XX
PP 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RBCG) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX

DR WPI: 1999-338008/28.
DR N-PSDB: X55771.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure: Page 49-50; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commusless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 434 AA;

Query Match 99.8%; Score 2276; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.1e-144;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIVAGGRTVTFPCETKGNPQPAVFWQKRGSSQNLFPNQPPQNSRCVSPGDLITNIQ 60

Db 1 qivaggrtvtfpcetkgnppavfwkrgssqnlfpnqppqnsrccvsgptdltitniq 60

Qy 61 RSDAGTYICQALTWAGSILAKAQLEVDVLTDRPPPIILQGPANQLAVDGTALLKCKAT 120

Db 61 rsdagtyicqaltwagsilakaglevtdvltdrpppilqgpanqlavdgtallkckat 120

Qy 121 GDLPLVISMWAKGFTFPGDRPATIQEGTGLQIKNLISDGTTCVATSSSGEASNAV 180

Db 121 gdlplvismwkgftfpgdrpatiqeggtglqiknlisdtgtytcvatsssgesav 180

Qy 181 LDVTESGATISKNDLSLPGPPSKPQVTDVTKNSVTLWSQPGTGLPASAYITIAFSQ 240

Db 181 ldvtesgatisknydsldlpgppskpqtvdtknsvtlwsqpgtgpilpasayileafsq 240

Qy 241 SVNSNQTVANHVHTLTITVRLRNPTIYLFWRANPKVSVTQXPKKNGSTWANVPL 300

Db 241 svnsnqtvanhvhtltitytrgipntilylfnwainpkvsvtgkpkngstwanvpl 300

Qy 301 PPPPVQLPGLTEHNAVQEGNGTDSWCPLPQVQTYLHQLEDELEDDORVPTFPV 360

Db 301 ppppvqlpglehnavqegngydsdswcplpqvtylqgledeleeddorvptfpv 360

Qy 361 RGVASSPAISFGQGTATLTSPREEMQMLQASPTSSQRPRTSPFSDTNSAALS 420

Db 361 rgvasspaifsggtatltspreemqmlqasptssqrprtpsfdsntsaals 420

Qy 421 QSQRPRTTKKHGG 434

Db 421 qsqrrprtckhgg 434

RESULT 2
Y08405
ID Y08405 standard; Protein; 434 AA.
XX
AC Y08405;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human partial ROBO2 protein.
XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
XX

PN W09920764-AL.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (RSCC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI: 1999-312615/26.
DR N-PSDB: X57254.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 72-73; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 434 AA;

Query Match 99.8%; Score 2276; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.1e-144;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIVAGGRTVTFPCETKGNPQPAVFWQKRGSSQNLFPNQPPQNSRCVSPGDLITNIQ 60

Db 1 qivaggrtvtfpcetkgnppavfwkrgssqnlfpnqppqnsrccvsgptdltitniq 60

Qy 61 RSDAGTYICQALTWAGSILAKAQLEVDVLTDRPPPIILQGPANQLAVDGTALLKCKAT 120

Db 61 rsdagtyicqaltwagsilakaglevtdvltdrpppilqgpanqlavdgtallkckat 120

Qy 121 GDLPLVISMWAKGFTFPGDRPATIQEGTGLQIKNLISDGTTCVATSSSGEASNAV 180

Db 121 gdlplvismwkgftfpgdrpatiqeggtglqiknlisdtgtytcvatsssgesav 180

Qy 181 LDVTESGATISKNDLSLPGPPSKPQVTDVTKNSVTLWSQPGTGLPASAYITIAFSQ 240

Db 181 ldvtesgatisknydsldlpgppskpqtvdtknsvtlwsqpgtgpilpasayileafsq 240

Qy 241 SVNSNQTVANHVHTLTITVRLRNPTIYLFWRANPKVSVTQXPKKNGSTWANVPL 300

Db 241 svnsnqtvanhvhtltitytrgipntilylfnwainpkvsvtgkpkngstwanvpl 300

Qy 301 PPPPVQLPGLTEHNAVQEGNGTDSWCPLPQVQTYLHQLEDELEDDORVPTFPV 360

Db 301 ppppvqlpglehnavqegngydsdswcplpqvtylqgledeleeddorvptfpv 360

Qy 361 RGVASSPAISFGQGTATLTSPREEMQMLQASPTSSQRPRTSPFSDTNSAALS 420

Db 361 rgvasspaifsggtatltspreemqmlqasptssqrprtpsfdsntsaals 420

Qy 421 QSQRPRTTKKHGG 434

Db 421 qsqrrprtckhgg 434

RESULT 3
Y08404
ID Y08404 standard; Protein; 1649 AA.
XX
AC Y08404;

XX
DT 24-JUL-1999 (first entry)
XX
DE Human ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
XX
PN W09920764-AL.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
20-OCT-1997; 97US-0062921.
XX
(RESC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X08404.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 65[71]; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1649 AA;

Query Match 40.1%; Score 914; DB 20; Length 1649;
Best Local Similarity 23.2%; Pred. No. 9e-53;
Matches 256; Conservative 54; Mismatches 118; Indels 682; Gaps 12;

QY 1 QIVAGQRTVTPCETKGNPQAVFVKQKGGQNLFPNQPPQNSRCVSPTGLTINIQ 60
360 QVVALGRTVTPCETKGNPQAVFVKQKGGQNLFPNQPPQNSRCVSPTGLTINIQ 419

QY 61 RDSAGTITQALTVAGSILAKAQLVETDVLDRPPPIILQSPANQTLAVDGTALLACKAT 120
DB 420 RDSAGTITQALTVAGSILAKAQLVETDVLDRPPPIILQSPANQTLAVDGTALLACKAT 479

QY 121 GDPLPISWNLKGGTTPFGDRPRATIQEQTGLQKNLRISDTGTTCVATSSSGEASNAV 180
DB 480 GSPPTLILVRKGVIVETGQSRILKGLVGLYKGLDGTGYTCLASTSPGEATWSAY 539

QY 181 LDVETSGATIS-KNYDLSQLPGPSKQVQVDTNKNVSLWQPGTPTLPSATYIIIAF 238
DB 540 LEVEGTVTPGTPRPTDNLIPSPKSPKQVQVDTNKNVSLWQPGTPTLPSATYIIIAF 599

QY 239 SQSVSNWSQVTVANVHTLTYVGLRPNITLYLPMVRAIN----- 277
DB 600 SHASGSSQVTVANVHTLTYVGLRPNITLYLPMVRAIN-----PK 279

QY 278 ----- 277
DB 660 TSQGVKHKVQRELGNVHLHNPVTLSSSIEVHWTVQGSQYIGQYKILYRPSGANHG 719

QY 278 -----PK 279
DB 720 ESDVLVFEVTPAKNSVPIDLRKVNYEIKARPFNEFGADSEIKFAKLEASPSP 779

QY 280 VSVTQXX----- 286
DB 780 QGVTVSKDQNGTALLVWQGGPQDQNGVMVGEYKVCWGLMETRYHINKTVGDSIFSWI 839

QY 287 -----PO----- 288
DB 840 PFLVPGIRYSVEVAASTGAGSVKSEPFILGLDAHGNVSPEDQSALQGISDVKQPAF 899

QY 289 -----KNG----- 292
DB 900 IAGIGACWILLMFSILWYLRHKRKNLTSTYAGIRKVPSTFTPTVYRQGEAWSSG 959

QY 293 -----STWAN----- 297
DB 960 GRPGLINSEPAAGVLDATWPTGNHNDSCSCTAGNDSNLTYSRPADCIANY 1019

QY 298 ----- 297
DB 1020 NNQLDKQTNLMPESTVYGVGLSKNKEMKTFSPNLDGFRVSPSQPTPYATTIGS 1079

QY 298 ----- 297
DB 1080 NLSNMMNGSGSDGKHWKPLGQKQEVAPGVNVEQNKINKYDRANDTVPTIYNGS 1139

QY 298 -----VP-----LPPPPVQPLPCTELEH 315
DB 1140 YDQGTGSGYNSDGSSTGSGGKKGARTKPKVGGGMWADLPPPAHPPHNSSEE 1199

QY 316 YAVEQKQNGYSDSGWCPPLPVQTYLHGLEDEL-EDDDRVPTPPVRGVAASP-AISFGQ 373
DB 1200 YNLVSDS-YDQEMCPVPVPMYIQQ-----DELEEDERGTPTPVGAASSPAVSYSH 1255

QY 374 QSTATLTGSPSPEDNQMLQAS----- 395
DB 1256 GSTATLTGSPSPEDNQMLQAS----- 1315

QY 396 ----- 395
DB 1316 LVSDMDTADPEEDEDMEVAKMQRILLRLGLEQTPASSVGLDSSVTSMINGWSA 1375

QY 396 -----XTSSQR 402
DB 1376 SEEDNLSSGRSVSSSDGSFFTDADFAGVAAAAYAGKVARRMQDAQGRHFRHSQC 1435

QY 403 PRPTSPFTSDNTSAALSQSRPRPTKXKNGG 434
DB 1436 PRPTSPFTSDNTSAALSQSRPRPTKXKNGG 1467

RESULT 4
Y13566
ID Y13566 standard; Protein; 1651 AA.
XX
AC Y13566;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
PN W09925833-AL.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RESC) UNIV CALIFORNIA.

Qy 298 ----- 297

```

FI      Domain          510..600
FT      /note= "cytokine receptor homology N-terminal
FM      domain"

```

XX
PN W09848051-A2.
XX
PD 28-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07714.
XX
PR 10-OCT-1997; 97US-0062017.
PR 18-APR-1997; 97US-0044746.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI Holtzman D, McCarthy SA;
XX
DR WPI; 1999-024021/02.
DR N-PSDB; V69278.
XX
PT New isolated human PTBMA-070 and 785 proteins - used to develop
PT products for the diagnosis and therapy of disorders involving
PT cellular processes, e.g. neuronal development.
XX
PS Claim 31; Fig 3; 127pp; English.
XX
CC This is the amino acid sequence of a novel human protein designated
CC 785, and also referred to as PMH-6D4 and FMBH-SD4. 785 cDNA (see
CC V69278) was identified in a human foetal brain cDNA library using a
CC screen designed to identify genes encoding proteins having a
CC functional signal sequence. 785 nucleic acids and polypeptides of
CC the invention are useful as modulating agents in regulating a
CC variety of cellular processes. They can be used for identifying
CC compounds which bind to or modulate the activity of the polypeptides
CC (claimed). They can also be used in screening assays, detection
CC assays (e.g. chromosomal mapping, tissue typing, forensic biology),
CC predictive medicine (e.g. diagnostic assays, prognostic assays),
CC monitoring clinical trials, and pharmacogenomics, and methods of
CC treatment (e.g. therapeutic and prophylactic) e.g. for neurological
CC disorders.
XX
SQ Sequence 753 AA;

Query Match 38.54; Score 877.5; DB 20; Length 753;
Best Local Similarity 51.14; Pred. No. 9.4e-51;
Matches 180; Conservative 48; Mismatches 93; Indels 31; Gaps 6;

Qy 1 QIVAGRTVTFPCETKGNPQANVFWKQSGNLLFPMPQPPNSRCSVSPGTGLTITNIQ 60
Db 324 qvvalgrtvtfqcaatgnppalfvrrsgnllfypgpgssrfsvsgtdlittvq 383

Qy 61 RSDAGTICQALTVAGSILAKAQLVTDVLRPPPIILQGANQTLVDGTALLKXAT 120
Db 384 rdsdygtylqclnvagsiitkaylevtdvldrpppvirgpgvngtvdvgtfvlscvat 443

Qy 121 GDLPLVSWLKBEGPTTFPGDRPRATIQEQTILQKNLRISDITGTYCVATSSGSEASNAV 180
Db 444 gspptllvrkgdlvstgdsrikqlengvlgiraykgldgtgrytclastpsgeatwsay 503

Qy 181 LDVTEGATIS--RWDLSDLPQPSKPGQVVDVKNVSLWSQPGTGLPSAYIIIAF 238
Db 504 lvevgtyvpgpprtpdnlipsapskpevdvsnrtvslwsgplnsagatpsyleaf 563

Qy 239 SQSVSNWSQTVANHVKTYTLTVRGLRPNTIYLFMVRIN----- 277
Db 564 shagsgswgtvraevktetsaikglkpnaiylfvranaayisdspsdgprvktgdvlp 623

Qy 278 PKVSYTQXKPKKNGSTWANNPLPPVPQPLGTELE-HYAVEQEE--MEY 325
Db 624 tsqgvdhkgvrelgn-avllhnptv-lsssslevhtwdqgsqygly 671

RESULT 6
Y13565
ID Y13565 standard; Protein; 1297 AA.

XX
AC Y13565;
XX
DT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo-Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1297 AA;

Query Match 24.68; Score 561.5; DB 20; Length 1297;
Best Local Similarity 37.24; Pred. No. 1.9e-29;
Matches 124; Conservative 48; Mismatches 114; Indels 47; Gaps 6;

Qy 1 QIVAGRTVTFPCETKGNPQANVFWKQSGNLLFPMPQPPNSRCSVSPGTGLTITNIQ 60
Db 325 qvvalgrtvtfqcaatgnppalfvrrsgnllfypgpgssrfsvsgtdlittvq 383

Qy 61 RSDAGTICQALTVAGSILAKAQL-----V 86
Db 384 rdsdygtylqclnvagsiitkaylevtdvldrpppvirgpgvngtvdvgtfvlscvat 443

Qy 87 TDVLTORPPPIILQGANQTLVDGTALLKXATGDLPLVSWLKBEGPTTFPGDRPRATIQ 146
Db 444 gspptllvrkgdlvstgdsrikqlengvlgiraykgldgtgrytclastpsgeatwsay 503

Qy 147 BQSTLQIKNLISDITGTYCVATSSGSEASNAVLDVTE--SGATISKNLSDLPQPPS 204
Db 504 tsqgvdhkgvrelgn-avllhnptv-lsssslevhtwdqgsqygly 671

Qy 205 KPQVTDVKNVSLWSQ-PGTGLPSAYIIIAFSQSVSNWSQTVANHVKTYTLTVRGL 263
Db 564 shagsgswgtvraevktetsaikglkpnaiylfvranaayisdspsdgprvktgdvlp 623

Qy 264 RPNITLYMVRIN-----PKVS--VTQXRP 287
Db 624 kphsyfvranaekgigtptsvssallvttsk 656

```

QY      1 QIVAGQGVITPFCETKGNPQAVPWQKESQNLPPQPPQOQNSKCSVSPGLDITLNIQ  60
Db      325 qvqvaggtatfectilvgqspayfvksegqgdlflfay-vsdagktkswptgtltlewr  383

QY      61 RSDAGYIQALVAVGSLAKRAQL-----V  86
Db      384 qvdeqyavcgmssagsslkkaaltfatfekgrvqkkskmgqkqknvgsilylisav  443

QY      87 TWLDVDRPPPIILQGPANLTLDVGLALKCATGDLPIVSIWAKRGPTPGDPRATIQ  146
Db      444 tgnptkppkptiehgqqlnvgsallpcqagsgkptgpiwrdglpdlidderishq  503

QY     147 EQGTILKLNRLSDTGTITTVATSSGSEASNAVDYTE--SGATISKNYDLQPPGPS  204
Db      504 stgslhiadlkpdtvyrtclaknedgestwaalstvedhtsnagfvrmppdnfnpsfpt  563

QY     205 KQGVTVKMSVYLSGQ--PCPTGLPASAVIIPASQGSVNSQTVANHTYLTILTVRLG  263
Db      564 qplivntdvtevelhmapetsagapptgilylgysspdlgvtfnlfpvyasteyrikgl  623

QY     264 RPNVILYLVRAIN-----PKVS--VYQXP  287

```

```

RESULT      8
YI3563
ID YI3563 standard; Protein; 1395 AA.
XX
AC
CC YI3563;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN WO9925833-Al.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (RREG ) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-33800R/28.
XX
NR P-SDB: X55767..
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 30-33; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
(commisuresless) polypeptide in contact with a cell expressing active
Robo (roundabout) on its surface. The method comprises modulating the
effective amount of Comm polypeptide in contact with the cell, where the
amount of expressed active Robo is specifically modulated inversely with
the modulation of the effective amount of Comm in contact with the cell.
The method is used to modulate the amount of active Robo expressed on a
cell. The method can be used to screen for agents that modulate Robo:Comm
interactions. This is particularly useful for modulating nerve cell
function.
XX
SQ Sequence 1395 AA;

Query Match          23.9%; Score 545; DB 20; Length 1395;
Best Local Similarity 40.9%; Pred. No. 2.6e+28;
Matches 115; Conservative 48; Mismatches 104; Indels 14; Gaps 7

QY    9 VTFPCETKGNPQPAPVFWKEGSSQNLLFPNQPPNSRCSSVSFGDLITINQRSDAGTY 68
Db    352 vqlpmasagpppsvfwkgeystlnfp---sshgrvyvaadgltqtdrqedegyyv 418

QY   69 COALTVAGSLAKAQLEVTVLDRPPEIILOGNPATLVNDGTALLKCAKCATDPLPVT 128
Db   419 csaifsvdsdvrvfvlqvssv-derpppliglpangatlkpgswatlpcratgnpsprik 477

QY  129 WLKEPTFFPGEDPRATIQEGTILQIKNLISDGTGYTCVATSSSGASGVAVLDVTEGA 188
Db  478 wfhdghavqagn-rysilggsslrvddlqlsdsgtytctasgergetswaatlvkeqgs 536

QY  189 T-ISKNLYLSDLPGPSPKPQVDVTKNSVTLGW-----OPGTGPTLPASAVITAFSGSW 242
Db  537 tshlradsptypappgtpkvnlvstrtslsrlwaksgekpgavg--piigtvyfspdl 594

QY  243 SMSWQTVAHVKTITLTIVGLRGENTIIYLFWRNPK-VSV 282

```

Qy	1	QIQAQGRVTFPFCTCEKPNQIPAVQKREQSQLNLPFQDQPMNSRCSVPSPDQ-LITIN	58
Db	306	qlwleqdevlefcqanqhrpilywvegnslillyp-rdgmenvitlpegrsvlsiar	364
Qy	59	IQRS DAGTI-COALTVAGSILAKLAQLEVDITLDRFPPIILQGFANQLADVGTALLCK	117
Db	365	faredsgvltcnalnagvsrvsrrvsv-dtqfelppileegpvnqnlpkvsvlwpk	423
Qy	118	KATGDPLPVIWMLKEQPTPGRD-PRATIQDGTQIQLKNL-RISDGTGTCVATSSSGBA	175
Db	424	rltqlpvpqsvyldipidipeqhernlsdgaqltsldlrhedeqltvcasnrgks	483
Qy	176	SWASVLDW---TESGATISKNYDLSDPGPPSKQVDTVKNSWTLWS-QPQTGPLTAS	231
Db	484	swyylrltdtpnnkvlfrapeltstlpggppkpmqvegnvsgvslwstrnkvggssvl	543
Qy	232	AGIIEELTDPGPNKIFAGNHWKTYTLTVGGLPPIILWENWILPKWSVYQAKPMKN	291
Db	544	gyvimeiknetdgwagvtrvqntfttqtlgpyvnyfllraensh-----	591
Qy	292	GSTWANNPLPPPPVPLP-CLEHLYAVBOO-----NGYDSNCSPPPLP	336

QY	1	QIQAQGVTVTFPCKTNGNPQAPVFWKESQMLLPFOQOPNSRCSVPFGD-LITIN	58
Db	306	qlvelgdevlfecagqghprtlvsvgegnssllipy-rdgmewtltegrvslsiar	364
QY	59	IQRSDAITGI-CQALLTVGSIILAAKQLEVTVLDRPPILLPGANPQTLAVDTALLK	117
Db	365	faresdskvltcnalnagvsvsrstvsv-dtqfelleppgvpnqtlpvksivlvp	423
QY	118	KATGDLQPLVLSMKEGTFFPGD-PRATIQEGTGLKLNK-RISDQGTITCVATSSGSA	175

Db	424	rtlgtppvpsvwydlgpldvqvehernrnsldageltisdlrhgedeglytvcvssrnrgks	483
QY	176	SHSAVLDT-----TSGATATSKNDLIDTGGPPSPKSPVITDTRKNSVTLG--PQGTGPLTAS	231
Db	484	swsylvrlndtpnplkfrefpelstlytppgppqmwqegensvltstrnkngvgssvl	543
QY	232	ATIIIAFSSQSSNSQWTVANHVVTLTVGLGRNPITYLEWVIRNPKVSVTQKPKQKN	291
Db	544	gyvlmfngkntdgvavvtrvgntfttqglpgvnyffirlanesh-----	591
QY	292	GSTWNVLPPLPPQPLP-STELERHAVQQQ-----NYGDSDSQSPPLPV	336
Db	592	-----glspsempsepiwtgtryfngsldlsearalsgdrvslsnavsvdtmsklwt	646
QY	337	Q-----TYLAQGLELEEDDRVPTPPGVGVAISSFGQOSTATLTPSPAE	385
Db	647	qinqqkyvgyfvyfyrq-----lplnvl-----npagvt	676
QY	386	EMQPMQASQFSTSPQSRPTSPFSTDSNALSQSRPQRPTKHHKG	434
Db	677	stnplqloststssssssssslstknleee--akrdetnqssqq	722

```

RESULT 12
R17126
ID R17126 standard; Protein; 1911 AA.
XX
AC R17126;
XX
DT 17-OCT-1995 (first entry)
XX
DE Human PTP-OB.
XX
KW PTP-OB; protein tyrosine phosphatase; osteoclast; differentiation;
KW osteoclast; osteoporosis; bone; cancer; osteosarcoma.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..29
FT /label= Sig_peptide
FT Modified-site 250
FT /label= N-glycosylation_site
FT Modified-site 721
FT /label= N-glycosylation_site
FT Modified-site 919
FT /label= N-glycosylation_site
FT Domain 1253..1277
FT /label= Extracellular_domain
PN
PW N09507935-A.
PD 23-MAR-1995.
XX
PP 09-SEP-1994; '94WO-US10166.
XX
PR 14-SEP-1993; '93US-0122032.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Rodan GA, Rutledge SJ, Schmidt A;
XX
NR WPI; 1995-131318/17.
XX
NR N-PSDB; Q86478.
XX
PT Protein tyrosine phosphatase protein PTP-OB specifically expressed
PT in bone cells - modulators of which are used to treat, e.g.
PT osteoporosis, and prevent and treat bone loss and cancer.
XX
PS Claim 1; Page 44-45; 63pp; English.
XX
CC PCR amplification of cDNA derived from human osteosarcoma
CC Sca-2/B10 using primers based on conserved regions of protein

```


control, and cellular behaviour such as motility and contact
inhibitions. In addition they may affect abnormal or potential

Search completed: January 22, 2001, 12:19:47
Job time: 1744 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:16 ; Search time 325.28 Seconds
(without alignments)
901595 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 QIVAGRTVFPCTKGNPQ.....TSAALSQSQRPPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Indexed: 195891 seqs, 5790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_66:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB	ID	Description
1	911	40.0	1612 2	T30805
2	911	40.0	1651 2	T14160
3	755.5	33.1	1344 2	T14316
4	579	25.4	1273 2	T42405
5	388	17.0	874 2	T29548
6	331.5	14.5	2222 2	T13924
7	326	14.3	1907 2	850893
8	325	14.3	1499 2	150212
9	323.5	14.2	1501 2	158148
10	323.5	14.2	1863 2	846217
11	317	13.9	1894 2	C54689
12	308.5	13.5	1897 1	TDBULK
13	307.5	13.5	1912 2	A56178
14	305.5	13.4	1898 2	846216
15	297.5	13.0	1232 2	T43027
16	296.5	13.0	1277 2	T30532
17	295.5	13.0	2029 1	TDFFLK
18	292.5	12.8	1262 1	848758
19	292.5	12.8	1496 1	A48758
20	282	12.4	1443 2	150600
21	276	12.1	1028 2	158164
22	274	12.0	1197 2	T30581
23	269	11.8	1028 2	A53449
24	268	11.8	1257 1	A41060
25	266.5	11.7	1427 2	151669
26	265.5	11.6	1260 1	805479
27	264.5	11.6	1256 2	T03096
28	262.5	11.5	1272 2	826180
29	261.5	11.5	1259 2	836126

30 258 11.3 1239 1 A32579 neuroglian - fruit
31 256.5 11.2 423 2 T29549 hypothetical prote
32 256 11.2 1040 2 A49356 transient axonal g
33 255.5 11.2 1375 2 T13822 frazzled gene prot
34 253 11.1 1018 2 J04211 axonin 1 precursor
35 252.5 11.1 1259 2 A34325 Bravo/Nr-CAM cell
36 251.5 11.0 1036 2 S22383 neural adhesion pr
37 251 11.0 761 1 IJWNG neural cell adhesi
38 251 11.0 1880 2 T18531 tractin - medicina
39 250.5 11.0 871 1 I48696 protein-tyrosine k
40 250.5 11.0 881 1 I48697 protein-tyrosine k
41 250.5 11.0 1268 1 A39640 neural cell adhesi
42 250 11.0 1091 2 A58532 glial cell membran
43 249 10.9 1447 2 A54100 tumor suppressor p
44 248.5 10.9 1437 2 T31093 probable protein-t
45 248 10.9 725 1 IJWNG neural cell adhesi

ALIGNMENTS

RESULT 1
T30805
dutt1 protein - mouse
N:Alternate names: transmembrane receptor protein Robol homolog
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30805
R:Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.
submitted to the EMBL Data Library, July 1998
A:Description: The mouse homologue of human DUTTL/H-robol gene: protein sequence and
A:Reference number: 120879
A:Accession: T30805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1612 <WU>
A:Cross-references: EMBL:J17793; NID:el329712; PID:el329713; PIDN:CAA76850.1
A:Experimental source: brain
C:Genetics:
A:Gene: dutt1
A:Map position: 16

Query Match 40.0%; Score 911; DB 2; Length 1612;
Best Local Similarity 22.8%; Pred. No. 1.7e+46;
Matches 254; Conservative 55; Mismatches 120; Indels 686; Gaps 10;

Qy 1 QIVAGRTVFPCTKGNPQAVFVKGQSSGNLFFNPQQPQNSRCVSPTGLDITNTIQ 60

Db 321 IVVAGRTVFPCTKGNPQAVFVKGQSSGNLFFNPQQPQNSRCVSPTGLDITNTIQ 380

Qy 61 RSDAGYIQQALTVAGSILAKAQLEVTDLTPPPILQPGPANTVAVDGLTLLSKCAT 120

Db 381 RSDAGYIQQALTVAGSILAKAQLEVTDLTPPPILQPGPANTVAVDGLTLLSKCAT 440

Qy 121 GDPLPWSLAKGPTPGPRDPRATQBOGTQLKRLNRISDTGTYTCVATSSSGSASW 180

Db 441 GSPAPTLAKRGVLSVQDSRIKQLSVGLQRYAKLGDTRGTYTCVATSSSGSASW 500

Qy 181 LDVTEGATIS - KNYLSDLPGPSKQPVDTKVNKLSWQPGTPTGLPASATIEAF 236

Db 501 LDVTEGATIS - KNYLSDLPGPSKQPVDTKVNKLSWQPGTPTGLPASATIEAF 560

Qy 239 SQSVSNNSQTVANEHKTITLVGLRPNTIPLWRAVIN - - - - - 277

Db 561 SHAGSSQWTAENKWTETPAIKGLKPNALYELVRAANAGISDPSQISDPVKTVQVP 620

Qy 278 - - - - - 277

Db 621 TSQGVDRHQVQRELGNVVLHNPITLSSSSVEVHWVQQSQYIQGYKILRPSGASHG 680

Qy 278 - - - - - PK 279

Db 681 ESEWLVFEVTRPTKNSVVIDLRKGVNVEIKARFFNEPGADSEIKFAPTLLEEPSAPP 740
Qy 280 VSVT 283
|||
Db 741 RSTVSKNDGNGTALTWTWQPPEDTQNGWQVYKWCMLGNETYHINKTVDGSTFSWV 800
Qy 284 283
Db 801 PSLVPGIRYSVEAASGAGPGVKSPEQFIQLDSHGPNVSPEDQVSLAQISDVVQKAP 860
Qy 284 283
Db 861 IAGIGAACWILMWFSINLIRHKKRNGLTSTYAGIRKVSPTPTPTVTYQGGEAVSSG 920
Qy 284 283
Db 921 GRPGLNISEPATQFWLADTWPNWNGSHNDCSINCCAGNGSNDNLTTSPADCIANY 980
Qy 284 283
Db 981 NNQLDNKQTNMLPESTVYGVGLNSKNEMKTPNSNKLKGRFVNSGQPTPIVATQLI 1040
284-QKQG- 288
|||
Db 1041 QANLNNNGGAGDSSEKHWKPPGQKQEVAPIQYINMEQNKLNKDTRANDTILPTIPIYN 1100
Qy 289 312
|||
Db 1101 QSTQDTGGSTNSDRGSSSTSGSQGHKGARTPAKPGQGNWADLLPPPPHPPHNS 1159
Qy 313 EHYAVEQENGYSDSNCWPLPVQTYLQGLEDEL-EEDDRVPTPPVGVASSP-AIS 370
|||
Db 1160 SEETNSVDES-IDQEMCPVPFAPMTLQQ---DELEQEEDEERGPTFPVGAASSPAVS 1215
Qy 371 QGQSTATLTPSPREEMQMLQASP 395
|||
Db 1216 TSGQSTATLTPSQBELQMLQDCPELGHMHPHPPDRRRQPVSPPPPPRISPHHTGYI 1275
Qy 396 395
Db 1276 GPLVSDMDTAPREEDDEADMEVAKMOTRLLRLGLEQTPASSVGLSSVTSVMINGW 1335
Qy 396-XTFS 399
|||
Db 1336 GSASEDNISSGRSVSSDGSFTTDAFAQAVAAAAEYAGLKVARQMDAAGRHHFAS 1395
Qy 400 SQRPRTSPFSTDSNTSAALSQSRPPRTKKHGG 434
|||
Db 1396 SQCPRTSPVSTDSNMSAVIQKAPAKQKQPG 1430

RESULT 2

T14160

transmembrane receptor protein Robol - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14160

R:Kidd, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodman, C.S.;
Cell 92, 205-215, 1998

A:Title: Roundabout controls axon crossing of the CNS midline and defines a novel subfam

A:Reference number: 217897; M0ID:98117249

A:Accession: T14160

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1651 <KID>

A:Cross-references: EMBL:AF041082; NID:2811215; PID:2811216; PIDN:AAC39960.1

C:Function:

A:Description: appears to function as the gatekeeper controlling midline crossing

C:Keywords: transmembrane protein

Query Match 40.0%; Score 911; DB 2; Length 1651;

Best Local Similarity 23.04%; Pred. No. 1.8e-46;

Matches 256; Conservative 57; Mismatches 117; Indels 684; Gaps 12:

Qy 1 QIVAGRTVTPFCETKGNPQAVFWKQSSGNLLFPNQPPQNSRCSVSGTDLITNIQ 60
|||
Db 360 QIVAGRTVTPFCETKGNPQAVFWKQSSGNLLFPNQPPQNSRCSVSGTDLITNIQ 419
Qy 61 RSDAGTYICQALTIVAGSILAKAQLVTDVITDRPPPILOQPANQTLVADGTALLKCAT 120
|||
Db 420 RSDAGTYICQALTIVAGSILAKAQLVTDVITDRPPPILOQPANQTLVADGTALLKCAT 479
Qy 121 GDPLFVSWLARGPTTPGRDPRATIQOGLTKNLRLSDTGTGTYTVAISSSGEASHAV 180
|||
Db 480 GSPVPTILMKDGLVSTQDSRAIKQLESCVLIIRYAKLGDTCRTCTASTPSGEATWSAY 539
Qy 181 LDVTESGATIS--KNYDLSLPGGSPKQVDTWKNSVTLWSQGTPTPLPASAIIEAF 238
|||
Db 540 IEVQEGVGVPPPTDPMILPSAPSKPEVTDVSKMTVILLQWNLGASPTSTIIIEAF 599
Qy 239 GQSVSNQTVANHWKTTILTVGLRPTNITLYMVRALN- 277
|||
Db 600 SHAGSSWQTVAEWKTETPAIKGLKPNALFLYLRANANIGISDPSQISDPKVTQDVP 659
Qy 278 277
Db 660 TTQGVHKGVRELGNVWHLANPTILSSSVVEVHTVDQSQPIQGYKILIRPSGASH 719
Qy 278-PK 279
Db 720 ESEWLVFEVTRPTKNSVVIDLRKGVNVEIKARFFNEPGADSEIKFAPTLLEEPSAPP 779
Qy 280 VSVTQK- 286
|||
Db 780 RSTVSKNDGNGTALTWTWQPPEDTQNGWQVYKWCMLGNETYHINKTVDGSTFSWV 839
Qy 287-PQ- 288
|||
Db 840 PFLVPGIRYSVEAASGAGPGVKSPEQFIQLDSHGPNVSPEDQVSLAQISDVVQKAP 899
Qy 289-KNG- 292
|||
Db 900 IAGIGAACWILMWFSINLIRHKKRNGLTSTYAGIRKVSPTPTPTVTYQGGEAVSSG 959
Qy 293-STWAN- 297
|||
Db 960 GRPGLNISEPATQFWLADTWPNWNGSHNDCSINCCAGNGSNDNLTTSPADCIANY 1019
Qy 298 297
Db 1020 NNQLDNKQTNMLPESTVYGVGLNSKNEMKTPNSNKLKGRFVNSGQPTPIVATQLI 1079
Qy 298-VP- 299
|||
Db 1080 QANLNNNGGAGDSSEKHWKPPGQKQEVAPIQYINMEQNKLNKDTRANDTILPTIPIYN 1139
Qy 300-LPPPPVQLPGLTE 313
|||
Db 1140 HSDQDTGGSTNSDRGSSSTSGSQGHKGARTPAKPGQGNWADLLPPPPHPPHNS 1199
Qy 314 EHYAVEQENGYSDSNCWPLPVQTYLQGLEDEED-DRVPTPPVGVASSP-AISP 371
|||
Db 1200 SEETNSVDES-IDQEMCPVPFAPMTLQQ---DELEQEEDEERGPTFPVGAASSPAVS 1255
Qy 372 QGQSTATLTPSPREEMQMLQASP 395
|||
Db 1256 TSGQSTATLTPSQBELQMLQDCPELGHMHPHPPDRRRQPVSPPPPPRISPHHTGYIS 1315
Qy 396 395
Db 1316 GPLVSDMDTAPREEDDEADMEVAKMOTRLLRLGLEQTPASSVGLSSVTSVMINGW 1375
Qy 396-XTFS 400
|||
Db 1376 GSASEDNISSGRSVSSDGSFTTDAFAQAVAAAAEYAGLKVARQMDAAGRHHFAS 1435

QY	86	VTDVLDRPPHLLQAGNANLAVCTALCKAKAGDPLVISMKEGTEFGPGRPATI	145
Db	20	VGTNTPAKPPTTEHQHQQTLLVGLSSALLPQASGKPTPGISLWRDGLPDITDSRISQ	79
QY	146	QEGQTQKNNRISDGTGYTCVATSSSGEASVADLTE--SGATISKNYDLSDLGGP	203
Db	80	TSKSLHIAADLKPDGTGYTCTIAKNEGDSWTSVLDYHDSNAQFNMPSNPSSP	139
QY	304	SKQGVTDKNSVLSVLS--PGTGPPLASALYIAFSGVSNSGVQVQVNAKHTLTVG	262
Db	140	TOPVIVNTDVEFLNWNAPPSAGGPTGYIYVSPSGDQWENPDIYASTETPRIG	193

RESULT 7
 S50893
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type sigma precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jan-2000
 C:Accession: S50893; S40281
 R:Wagner, J.; Boerboom, D.; Tremblay, M.L.
 Eur. J. Biochem. 226, 773-782, 1994
 A>Title: Molecular cloning and tissue-specific RNA processing of a murine receptor-type
 A:Reference number: s50893; MUID:95112841
 A:Accession: S50893
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1907 <GAG>
 A:Cross-references: EMBL:X82288; NID:s587483; PIDN:CAA57732.1; PID:s587484

QY 1 QIQAQGRVIFPFCETGNQPAVKEQSGKNLFPQOPQMSRCSVSPGDDLTINIQ 60
Db 144 KVVERTRATMLCAASGNPDEITFWKD----FLPVPDSASNGRIKLGSAQIQUES 198

QY 61 RSDAQITCYQALITVAG--SILAAQLEVTVDTRAPPQIILQSGPANQTLAVDGTALLK 117
Db 199 ETQQKELCVATNSAGVTRSSPANLYVRVRY--APRFSL--PMSHILPGGNWITC 233

QY 118 KATGDPLVSLWKEGTFPGRDPRATIQEQETQLIKMLRISDTGTTVCATSSGEASW 177
Db 254 VAGVSPMPVWQKQAGDQLTPEDDMPV---GRNVLETDVKDSANTCVAMSSLG-- 305

QY 178 SAVLDVTEGATSKNTDLSLPPSPKQVTDVTKNSVTLWSQPGPTLPGTASAIIEA 237
Db 306 ----VIEAVQIT---VKSGLPAKPQTVIENTATSIITVWSDGNP--PVSYVIEY 354

QY 238 FQSQSVNSQWTVAHEVKTITLYTVRGLRNTIYFLMVRINPKVSVTXQKPKNGNS-TWA 296
Db 355 KSKSQDGPQ-IREDTITTRISIGLSPNSNETIEMVSAW--SIGQPPSSVWRTIGE 410

QY 297 NVPLFP--PVQP--LPGETLHYVAVBQONGTDSQKCP 334
Db 411 QAPASAPRNQARMLSATTINQVEEVPVNGLLRGYV-TYME-----PEH 457

QY 335 PVQITLHQGLD-----ELEEDD-----DRVPTPVRGVASSP-- 367
Db 458 PVGNWQKHWDSLLTIVGSLDEETITVRLAFTSGDGLSDPQWTKQVQPGVQPN 517

QY 368 -----AISE-----GQSQATLTPSPR--EMQP-- 389
Db 518 LRAEAKSETSIGLWSAPQBSVYIKYLLFROGDGRGVERTDPTATVVEDLKPIEY 577

QY 390 ---MLQASP---XFTS---SQRPRTSP---FSTDSMTSAAQSQRPRTKHKHG 433
Db 578 AFRLAARSPQGLAFTVAHVHOTOIQAQSPAPDQVKCSLRSALTALVSWRRPPEHNG 636

```

RESULT      8
I50212
protein-tyrosine-phosphatase (BC 3.1.3.48) - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jan-2000
C:Accession: I50212
R:Stocker, A.W.
  Mech. Dev. 46, 201-217, 1994
A:Title: Isoforms of a pcvsl cell adhesion molecule-like protein tyrosine phosphatase

```


RESULT 10
S46217
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat
N:Alternate names: leukocyte common antigen-related phosphatase

us-09-540-245a-19.rpr

[illegible]

SSUBSU 11
 C54689
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form B precursor
 N:Alternate names: MPTP delta type B/C
 N:Contents: protein-tyrosine phosphatase, receptor type delta, splice form C
 C:Species: Mus musculus (house mouse)
 C:Date: 25-Apr-1995 #sequence_revision 19-May-1995 #text_change 12-Feb-1999
 C:Accession: C54689; B54689
 R:Mizuno, K., Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Takura, H.
 Moll. Cell. Biol. 13, 5513-5523, 1993
 A:Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in special
 A:Reference number: A54689; MUID:93360986
 A:Accession: C54689
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1894 <MID>
 A:Experimental source: brain; splice form B
 A:Note: sequence inconsistent with nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIIN:137486, NCBIPI:137487)
 A:Accession: B54689
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-352, H',354-535, F',537-601,1002-1894 <MID>
 A:Experimental source: brain; splice form C
 A:Note: sequence inconsistent with nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIIN:136527, NCBIPI:136530)
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog
 ogy
 C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester
 P:45-107/Domain: immunoglobulin homology <IMID>
 P:245-399/Domain: immunoglobulin homology <IM2D>
 F:317-399/Domain: fibronectin type III repeat homology <FN3A>
 F:1278-1894/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 P:1652-1874/Domain: protein-tyrosine-phosphatase homology <PTP2>
 P:1536/Active site: Cys (phosphocysteine intermediate) #status predicted
 P:1542/Binding site: substrate phosphate (Arg) #status predicted
 P:1826/Active site: Cys (phosphocysteine intermediate) #status predicted
 P:1832/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.9%; Score 317; DB 2; Length 1894;
Best Local Similarity 23.4%; Pred. No. 4e-11;
Matches 128: Conservative 63; Mismatches 188; Indels 168; Gaps 22;

QY	1	QIQAQGVTFPFCTCKGNKPQAVFWQKESQNLFPPOQPOPNRSVCSPSPGDLITTNIG	60
Db	144	KVVERTKATMLCAASGNPDPEITWFD----PLVPQVPSASNGRIKLSRGAQIESSE	198
QY	61	RSQGTITICQALIVAG----SILAAQAEIVDTVDRPPIIIGQNPANITAVDTALLK	117
Db	199	ETDQGEKCVKAVNSAGVRYSPANLTVYVRVY----APPSIL-PMSHEDMGGGNWITC	253
QY	118	KATGDPLEVISLWKEGPTTPGDRPRATIQEGTGLIKNLISITDGTITCVATSSGSAW	177
Db	254	VAVGSPMPTVYKMGQAEDLTPEDMFY----GRNVELTDVKDSANTVCAMSSLG--	305
QY	178	SAVLDVETSGATISKYNDLSLDPGPPSKPQVTDVYKSVTLSDWQDPTLPASAAIIEA	237
Db	306	-----VIEAVQIT----VKSLPKAPETPVVTVENATISITVWDGSDNP--PVSTIVIEY	354
QY	238	FSQSVNSMCTVAHVHKTITLTITVGLRPNITVFWIRAFKPVYSVQKKPKNNGS-VA	296
Db	355	KSKSGDQGTQ-IREKIDITTRTISIGLSPNSEYETIWSAVN----SIQGPPESSVTWRTGE	410
QY	297	NVLPPLPP-----PQVQ--LPQTELEHYAVQZQENGSDSKSCPK	334

QY 1 QIQAQGGTTPFCETKGNPQPAVQWQEGSNLPLPNPQDQPNRSCS---VSPPTGDLIT 57
Db 142 KVVERTATATMLCAASGNPPETITFKD----FLPVTSTNNNGRIKLQSRSEGAQIE 196
QY 58 NIGRSAGGYIQCALVAGS-LIAKAQELVTDLDRPPPIILQAPQNTLAVDTALK 116
Db 197 QSEESDQKICEYVANSNATRYSAAPALVYR---RVPPRFSPITPTNHEIMPQGSVNT 251
QY 117 KRATQNPPLVVISLKGFFPGPRPRATIQEQGTILKINRIASDTGTYTCVATSSGEAS 176
Db 252 CVAGSPMPYKVMIGADLTPEDMDPI-GRNLELDNVR-QSANTCYAMSLG--- 304
QY 177 NSVALVDTEGATISKNYDLSLDLPGPPSPQVDTVKNSTVLSMQQPTGTLPSAATIE 236

leukocyte antigen-related protein precursor - human
 N:Alternate names: leukocyte common antigen homolog
 N:Contains: protein-tyrosine-phosphatase (EC 3.1.3.48)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1991 [sequence_revision](#) 31-Dec-1991 [text_change](#) 22-Jun-1999
 C:Accession: S03841; J00051
 R:Streuli, W.; Krueger, M.X.; Hall, L.R.; Schlossman, S.F.; Saito, H.
 J. Exp. Med. 168, 1523-1530, 1988
 A:Title: A new member of the immunoglobulin superfamily that has a cytoplasmic region h
 A:Reference number: J00051; MUID:89035978
 A:Accession: S03841
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1897 <STR>
 A:Cross-references: EMBL:Y00815; NID:934266; PIDN:CAA68754.1; PID:934267
 C:Genetics:
 A:Gene: GDB:PTPRF; LAR
 Cross-references: GDB:120138; OMIM:179590
 Map position: lp34-lp34
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;
 ogy
 C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
 F:1-16/Domain: signal sequence [*status predicted](#) <SIG>
 F:17-1897/Product: leukocyte antigen-related protein [*status predicted](#) <NAT>
 F:17-1250/Domain: extracellular [*status predicted](#) <EXT>
 F:37-99/Domain: immunoglobulin homology <IM1>
 F:139-199/Domain: immunoglobulin homology <IM2>
 F:236-290/Domain: immunoglobulin homology <IM3>
 F:308-390/Domain: immunoglobulin type III repeat homology <FN3A>
 F:403-489/Domain: fibronectin type III repeat homology <FN3B>
 F:501-583/Domain: fibronectin type III repeat homology <FN3C>
 F:596-685/Domain: fibronectin type III repeat homology <FN3D>
 F:698-798/Domain: fibronectin type III repeat homology [*status atypical](#) <FN3E>
 F:810-983/Domain: fibronectin type III repeat homology <FN3F>
 F:905-989/Domain: fibronectin type III repeat homology <FN3G>
 F:1001-1078/Domain: fibronectin type III repeat homology <FN3B>
 F:1251-1274/Domain: transmembrane [*status predicted](#) <TM>
 F:1275-1897/Domain: intracellular [*status predicted](#) <INT>
 F:1285-1897/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F:1365-1586/Domain: protein-tyrosine-phosphatase homology <PTP1>
 F:1654-1877/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F:44-97,146-197,243-288/Disulfide bonds: [*status predicted](#)
 F:107,240,285,711,956/Binding site: carbohydrate (Asn) (covalent) [*status predicted](#)
 F:1538/Active site: Cys (phosphocysteine intermediate) [*status predicted](#)

F:1544/Binding site: substrate phosphate (Arg) #status predicted
F:1829/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1835/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.5%; Score 308.5; DB 1; Length 1897;
Best Local Similarity 28.7%; Pred. No. 1.3e-10;
Matches 118; Conservative 47; Mismatches 165; Indels 81; Gaps 17;

Qy	1	QIVAGQGRPPFCPTCKRQPPAVFQKSGQNLFFPQQPQMSRCSVSPGDLITNIQ	60
Db	134	KVVEKARTATLCAAGSNPDPEISWFD----FLVPDPATSNGRKILRSALQESSE	188
Qy	61	RSDAQITQICALTIVGS-ILAKAQLEVDVLDRPPPIILQSPANQITAVDGITALAKA	110
Db	189	ESDQKTCVATNSAGRTISAPANTLYR---VRRVAFRIPPSQGVPMGSGSWLTVCA	245
Qy	120	TGDPPLFWLVLG-----KEGTFPGKPRPATIQBGQTIQKMLRISDTGTTVCATWS	172
Db	246	VGAPPIVPMNGAEELTKEDENFVGRN-----VLELSH-VYRSANICVALSIL	294
Qy	173	GEASNSAVLDYTESGATIKNVDLSLQPPSPKQPIVDTVTKNSVTLSPQGPPTITPASA	232
Db	295	G-----MIEATQV-----VVALPKPIDLVTETTATSLVWNGS--NSEPVY	339
Qy	233	YIIEAFQSVSNQWTANWHKTYLTVVRLGAPNTIYLFMRVIANPKVSVTQXAKPKN-N	291
Db	340	YGIQIRAASTGPGFQVY-DGVAATIRYSIGLSFPFSEAFRLAVN---SIGCRPPSEAR	395
Qy	292	GSTNANVLPPLP-VPVP-----LPGTELEHYEAVQEGN-----TDSQSWCP--	333
Db	396	ARTGEQAPSPSPRRVQARMLSASTMLVQNPQPEPENGVLRGYEVITPTDPSRRNANWH	455
Qy	334	-----LPVITVYHQGLEDEEDDORVPTPPVR---GVASSPA	368
Db	456	NTDAGLLTVVSLPGITLTVSLVIAFTAVGQGPSPPTIOVKTQOQVAPAPA	506

AS56178 13
 AS56178
 protein:tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - human
 N:Alternate names: protein:tyrosine-phosphatase BPTP-2
 C:Species: Homo sapiens (man)
 C:Date: 03-Oct-1995 sequence_revision 03-Oct-1995 #text_change 21-Jan-2000
 C:Accession: AS56178; S12052; B44929
 R:Pulido, R.; Krueger, N.X.; Serra-Pages, C.; Saito, H.; Streuli, M.
 J. Biol. Chem. 270, 5722-5728, 1995
 A:Title: Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta isoforms. /
 A:Reference number: AS56178; MUID:95204468
 A:Accession: AS56178
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1912 <PUL>
 A:Cross-references: GB:J38929; NID:G755652; PIDN:AAC41749.1; PID:G755653
 R:Krueger, N.X.; Streuli, M.; Saito, H.
 EMBO J. 9, 3241-3252 / 1990
 A:Title: Structural variation and evolution of human receptor-like protein tyrosine p
 A:Reference number: S12049; MUID:91006018
 A:Accession: S12052
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 390-1912 <KRÜ>
 A:Cross-references: GB:X54133; NID:G35789; PIDN:CAA38068.1; PID:G35790
 A:Note: the sequence from Flg. 5B is inconsistent with that from Flg. 5A in having 56
 R:Adachi, M.; Sekiya, M.; Arimura, Y.; Takekawa, M.; Itoh, F.; Hinoda, Y.; Imai, K.;
 Cancer Res. 52, 737-740, 1992
 A:Title: Protein-tyrosine phosphatase expression in pre-B cell NALM-6.
 A:Reference number: A44929; MUID:92119637
 A:Accession: B44929
 A:Molecule type: mRNA
 A:Residues: 1756-1804, 'C', 1806-1845 <ADA>
 A:Cross-references: GB:878086; NID:A243545; PIDN:AB21147.1; PID:G243546

Page 8

[illegible]

Db 256 VGMPTVTVKMGAEELTKEDMPVGRN-----VLELEN-VMSANTTCAVSISS 304
Qy 173 GNASAVLVDSGATISKNYDLSLGGPPSKQVDTVKNSVLSWQPGTGLTASA 232
Db 305 G-----MTEATQVT-----VKALPKPIDLVVETATSVTLWDSG--NTPVVSF 349
Qy 233 YIIAFSQSWSNSQVTVANHVKTILYTVGLRPNTIYLVKRVAINKYSVIXKPKQN-N 291
Db 350 YGIQTRAGTDGPPQEV-DGVASTRYSIGLSPFSEATFVLAVN---SIGRGPPSEAVR 405
Qy 292 GSTWANVPLPPP-PVQP---LPGTELEHYAVEQENG-----YDSDS-----W--- 330
Db 406 ARTGEOAPSSPPRRVQARMLSASTMLVQWEPPEPNGLVRGTVTTPDSRRPLSAHKKH 465
Qy 331 -----CPPLP-VQTLHQGLEDE-----LEEDD 352
Db 466 NYDAGLLTVGSLRGITYSRLVLAFTAVGDSPPPTIQVTKQGVPAQPADPQAKESD 525
Qy 353 DRVP---TPPVKGVASSPAISF-----GQSTATLTPSPR---HEMOP----- 389
Db 526 TRIQLSWLLPQERIIRKLYLWAAEDBQKHVTFDPTSSVYLEDLAPOTLYHFQLAAR 585
Qy 390 -----MLQASPIXTSSQRPRTSFSTDSVNSAALSQSRPRTKKEHG 433
Db 586 SDLGVGVPTTVIETACTAGTSPSAPPQKVICVSTGSTR---VRVSWPPPADSRNG 637

RESULT 15

T43027

neural cell adhesion molecule L1 - goldfish

N:Alternate names: E587 antigen

C:Species: Carassius auratus (goldfish)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000

C:Accession: T43027

R:Giordano, S.; Leessing, U.; Lottspeich, F.; Stuermer, C.A.O.

submitted to the EMBL Data Library, April 1996

A:Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule exp

A:Reference number: 122294

A:Accession: T43027

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1232 <GID>

A:Cross-references: EMBL:U55211; NID:g1305526; PID:g1305527; PIDN:AAA99159.1

C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology;

C:Keywords: cell adhesion; membrane protein

Query Match 13.04; Score 297.5; DB 2; Length 1232;
Best Local Similarity 24.94; Pred. No. 3.4e-10;
Matches 111; Conservative 67; Mismatches 184; Indels 83; Gaps 17;

Qy 5 QGRVTTFPCETKGNPQPAVFWKQESQNLFPNQPPQNSRCSVPGDITINIRSDA 64
Db 409 EQQTLLQCRFTGSPQPKVDQITNS-----GPAANAKMSQTSQGNLQISDVSEDS 461
Qy 65 GTTICQALTVASILAQAQLEVTVDLDRPPPI-ILQGPANQTLAVDGTALLCKATGDP 123
Db 462 SMTCSSVSTNSMSISAEVLVLMRTRIVDPQDLRLVWG-----DDAVLQCRITVDH 512
Qy 124 L---PVISMLKE--GTFPGDRPRATIQDQGLIQLNKLISDTGTITCVATSSSGASWS 178
Db 513 MLQPTIQWKKDKHKITSSANDKTYRSPDGLKLTVDQMEDSGIYSC-----EIS 563
Qy 179 AVLD-VTESGATISKNYDLSLGGPPSKQVDTVKNSVLSWQPGTGLTASAYIIEA 237
Db 564 YLDSVSATGSIV-----VLDKPGSPHLELSKKERSVTLSSWPAENNSPISSEVIER 618
Qy 238 FQSQ--VSNSQTVANEVKTILYTVGLRPNTIYLVKRVAINKYSVIXKPKQNKGSTW 295
Db 619 KEKQNPQKGGWEETRRVQDITHEILHLQPYSTYHFRVGVN-GIGMSEPPSPSEYST- 676
Qy 296 ANVPLPPPPVQP-----LPGTELEHYAVEQENGIDSQWCPPLVQITLH 341

Db 677 ---PAAKDMPENVTISVSDNSLVITWQELS---QRFQNG-----PGFKIKITWR 722
Qy 342 QGLEDELEEDDORVPTPPVGVASSPAISFGQGSTATLTPSPREMQ-----PMLQA 393
Db 723 QCGDSHWMSSEASNPFFIVGPGTIFPQIKVQAVNVLGAGPEPDAGIETSGEDLP-LEA 781
Qy 394 SPXTSSQSRPPT-----SPFSYDS 413
Db 782 PSSVAVSELNKTTILVKNVSPVSTKS 806

Search completed: January 22, 2001, 12:27:28

Job time: 2125 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:29:42 ; Search time 162.41 Seconds
(without alignments)
86.298 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 Q1VAQGRVTPCTKGNPQ.....TSAALSQSGRPPTKKKGG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
	1	309.5	13.6	837	1	NCM2_HUMAN	Q15394 homo sapien
	2	308.5	13.5	1897	1	PTPF_HUMAN	P10586 homo sapien
	3	307.5	13.5	1912	1	PTPD_HUMAN	P23468 homo sapien
	4	295.5	13.0	2029	1	LAR_DROME	P16621 drosophila
	5	292.5	12.8	837	1	NCM2_MOUSE	O35136 mus musculus
	6	284.5	12.5	1377	1	NEOL_RAT	P97603 rattus norv
	7	282	12.4	1443	1	NEOL_CHICK	Q90610 gallus gall
	8	279	12.2	1493	1	NEOL_MOUSE	P97798 mus musculus
	9	268	11.8	1257	1	CAML_HUMAN	P32004 homo sapien
	10	267.5	11.7	1461	1	NEOL_HUMAN	Q92859 homo sapien
	11	265.5	11.6	1260	1	CAML_MOUSE	P11627 mus musculus
	12	261.5	11.5	1259	1	CAML_RAT	Q05695 rattus norv
	13	259	11.4	1447	1	DOC_HUMAN	P70711 mus musculus
	14	258	11.3	1239	1	NRG_DROME	P20241 drosophila
	15	256	11.2	1040	1	AXOL_HUMAN	Q02246 homo sapien
	16	251.5	11.0	1036	1	AXOL_CHICK	P28685 gallus gall
	17	251	11.0	761	1	NCA2_HUMAN	P13592 homo sapien
	18	250.5	11.0	1284	1	NRCA_CHICK	P35331 gallus gall
	19	249	10.9	1447	1	DOC_HUMAN	P43146 homo sapien
	20	248	10.9	725	1	NCA2_MOUSE	P13594 mus musculus
	21	247.5	10.9	1018	1	CONF_HUMAN	Q12860 homo sapien
	22	246.5	10.8	1115	1	NCA1_MOUSE	P13595 mus musculus
	23	245	10.7	1040	1	AXOL_RAT	P22063 rattus norv
	24	244	10.7	1010	1	CONF_CHICK	P14781 gallus gall
	25	243	10.7	2012	1	DSCA_HUMAN	O60469 homo sapien
	26	241.5	10.6	1020	1	CONF_MOUSE	P12960 mus musculus
	27	238.5	10.5	853	1	NCA1_BOVIN	P18336 bos taurus
	28	238.5	10.5	858	1	NCA1_RAT	P13596 rattus norv
	29	236.5	10.4	848	1	NCA1_HUMAN	P13591 homo sapien
	30	235.5	10.3	1906	1	KWLS_CHICK	P11799 gallus gall
	31	223	9.8	1266	1	NCA2_CHICK	Q03696 gallus gall
	32	222.5	9.8	898	1	FAS2_SCHAM	P22548 schistosom
	33	222.5	9.8	1142	1	MYPP_HUMAN	Q14324 homo sapien

34	222	9.7	1913	1	KWLS_HUMAN	Q15746 homo sapien
35	220	9.6	1051	1	PK77_CHICK	Q91048 gallus gall
36	218	9.6	811	1	FS22_DROME	P34083 drosophila
37	218	9.6	873	1	FS21_DROME	P34082 drosophila
38	215	9.4	1091	1	NCA1_HUMAN	P13590 gallus gall
39	215	9.4	4393	1	PGCM_HUMAN	P98160 homo sapien
40	214.5	9.4	1070	1	PK77_HUMAN	Q13308 homo sapien
41	213.5	9.4	819	1	FGRL1_CHICK	P21804 gallus gall
42	210.5	9.2	822	1	FGRL1_HUMAN	P11362 homo sapien
43	210.5	9.2	822	1	FGRL1_MOUSE	P16092 mus musculus
44	210.5	9.2	822	1	FGRL1_RAT	Q04589 rattus norv
45	210.5	9.2	3707	1	PGCM_MOUSE	Q05793 mus musculus

ALIGNMENTS

```

RESULT 1
NCM2_HUMAN
ID NCM2_HUMAN STANDARD; PRI: 837 AA.
AC Q15394;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE 2 PRECURSOR (N-CAM 2).
GN NCM2 OR NCAM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=9736993; PubMed=9226371;
RA Peoloni-Giacobino A., Chen H., Antonarakis S.E.;
RT "Cloning of a novel human neural cell adhesion molecule gene (NCAM2)
RT that maps to chromosome region 21q21 and is potentially involved in
RT Down syndrome."
RL Genomics 43:43-51(1997).
CC -1- FUNCTION: MAY PLAY IMPORTANT ROLES IN SELECTIVE FASCICULATION AND
CC ZONE-TO-ZONE PROJECTION OF THE PRIMARY OLFACTORY AXONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED MOST STRONGLY IN ADULT AND FETAL
CC BRAIN.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC OR SEND AN EMAIL TO license@isb-sib.ch).
CC -----
DR ENBL: U75330; AAB80803.1; .
DR MIM: 602040; .
DR INTERPRO: IP0001777; .
DR INTERPRO: IP0003006; .
DR PFAM: PF00041; fn3; 2.
DR PFAM: PF00047; fn5; 5.
KW Cell adhesion; Transmembrane; Glycoprotein; Repeat;
KW Immunoglobulin domain; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 837 NEURAL CELL ADHESION MOLECULE 2.
FT DOMAIN 20 697 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 698 718 POTENTIAL.
FT DOMAIN 719 837 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 100 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 129 193 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 225 288 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 315 387 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 415 481 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 482 581 FIBRONECTIN TYPE-III.

```

J. Exp. Med. 168:1523-1530(1988).
 [2]
 RX MUTAGENESIS.
 RP MEDLINE=90046860; PubMed=2554325;
 RA Streuli M., Krueger W.X., Tsai A.Y.M., Saito H.;
 RT "A family of receptor-linked protein tyrosine phosphatases in humans
 RL and Drosophila."
 RT Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
 RP [3]
 RX MUTAGENESIS.
 RA MEDLINE=90315093; PubMed=1695146;
 RA Streuli M., Krueger W.X., Thai T., Tang M., Saito H.;
 RT "Distinct functional roles of the two intracellular phosphatase like
 RT domains of the receptor-linked protein tyrosine phosphatases LCA and
 RL LAR."
 RL EMBO J. 9:2399-2407(1990).
 CC -1- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
 CC IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
 CC (PTPASE).
 CC -1- FUNCTION: THE FIRST PTPASE DOMAIN HAS ENZYMATIC ACTIVITY, WHILE
 CC THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE
 CC FIRST ONE.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collabora-
 CC tion between the Swiss Institute of Bioinformatics and the EMBL outstand-
 CC ing European Bioinformatics Institute. There are no restrictions of
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commer-
 CC cial entities requires a license agreement (See <http://www.isb-sib.ch/annual>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Y00815; CAA68754.1; -.
 DR PIR: S03841; TDHULC.
 DR HSP: P18052; TYPO.
 DR MIM: 179590; -.
 DR INTERPRO: IPR000242; -.
 DR INTERPRO: IPR000387; -.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFM: PFM00102; Y_phosphatase; 2.
 DR PFM: PFM00044; fn3; 7.
 DR PFM: PFM00047; lg; 3.
 DR PRINTS: PR000174; ENTPE111.
 DR PRINTS: PR00700; PRTYTPHPTASE.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE: PS00506; TYR_PHOSPHATASE_2; 2.
 DR PROSITE: PS00505; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
 KW Cell adhesion; Immunoglobulin domain; Duplication.
 FT SIGNAL 1..16 POTENTIAL.
 FT CHAIN 17..1897 LAR PROTEIN.
 FT DOMAIN 17..1250 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1251..1274 POTENTIAL.
 FT DOMAIN 1275..1897 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1360..1606 PROTEIN-TYROSINE PHOSPHATASE.
 FT DOMAIN 1649..1897 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 1538..1538 BY SIMILARITY.
 FT ACT_SITE 1829..1829 BY SIMILARITY.
 FT MUTAGEN 1538..1538 C>S: LOSS OF ACTIVITY.
 FT CARBOHYD 107..107 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 240..240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 285..285 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 711..711 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 956..956 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1897 AA; 211844 MW; 439850FD5C031FF CRC64;

Db 190 GALQIEQSEESDQGYECVATNSAGTRYSAKANLVRELREVRVPPRFSIPPTNHEIMP 249

PT	SIGNAL	1	32	
PT	CRAIN	33	2029	PROTEIN-TYROSINE PHOSPHATASE DLAR.
PT	DOMAIN	33	1377	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	1378	1402	POTENTIAL.
PT	DOMAIN	1403	2029	CYTOPLASMIC (POTENTIAL).
PT	DOMAIN	50	118	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	154	216	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	249	308	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	320	417	FIBRONECTIN TYPE-III.
PT	DOMAIN	418	512	FIBRONECTIN TYPE-III.
PT	DOMAIN	513	607	FIBRONECTIN TYPE-III.
PT	DOMAIN	608	706	FIBRONECTIN TYPE-III.
PT	DOMAIN	707	809	FIBRONECTIN TYPE-III.
PT	DOMAIN	810	906	FIBRONECTIN TYPE-III.
PT	DOMAIN	907	1007	FIBRONECTIN TYPE-III.
PT	DOMAIN	1008	1102	FIBRONECTIN TYPE-III.
PT	DOMAIN	1103	1207	FIBRONECTIN TYPE-III.
PT	DOMAIN	1492	1738	PROTEIN-TYROSINE PHOSPHATASE.
PT	DOMAIN	1781	2029	PROTEIN-TYROSINE PHOSPHATASE.
PT	ACT_SITE	1670	1670	BY SIMILARITY.
PT	ACT_SITE	1961	1961	BY SIMILARITY.
PT	DISULFID	57	111	POTENTIAL.
PT	DISULFID	1651	2029	POTENTIAL.
PT	DISULFID	256	301	POTENTIAL.
PT	CARBOHYD	176	176	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	253	253	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	258	258	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	553	553	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	616	616	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	666	666	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	721	721	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	774	774	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	915	915	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	962	962	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	1183	1183	N-LINKED (GLCNAc . .) (POTENTIAL).
PT	CARBOHYD	1304	1304	N-LINKED (GLCNAc . .) (POTENTIAL).
SO	SEQUENCE	2029	AA: 229027 MW: 536kDa	Q9A3D9 PRO CRF64

13.0%: Score 295.5: DB 1: Length 2029:

Best Local Similarity 23.34; Pred. No. 2.3e-10;
Matches 127; Conservative 66; Mismatches 180; Indels 173; Gaps 21;

Qy 1 QIVAGRTVTPPCETKGNPQAVKQESQNLFPNQPPQNSRCVS-SPGIDLTITNI 59
Db 45 QGVGVGVGASVTCARCDPPSPVWRKNG-----KKVSGTQSRITVLEQPGISILRI 97
Qy # 60 Q-----RSDAGTITQCALVAGSILAKAQLVDTVLDTRPP-----PIIQGPANQVLAVDG 111
Db 98 EPVARGDAPFECVACVGVGDAVSADATLITIE--GDKTPAGFPVITQGPOTVRVIEGH 155
Qy 112 TALLKCATGDPGLPVISWKEGTFPPGDRPATIQGQTLQIKNLISDTGTTCTVATSS 171
Db 156 TVLMTCKAIGNPTPIITWIKNQTKVMSNPVSLKD-GFLQIENSREDDQKTCVARN 214
Qy 172 SG-EASNSAVLDV-----TESGATISKNYDLS-----197
Db 215 MTEHSKATNLVYKVRVPPPTSRPETTISEVYLSGNLMSIAVGSMPHVKNMGSGED 274
198 -----DLPGPSPKSPVDVT 212
Db 275 LTPENEMPIGRNVLLQINQESANYTICIAASTLQGISVSVKVSQSLPATIDVTSVET 334
Qy 213 KNSVILNQSGPTGTLASAYILRAFSQSVNSQWVAVHVTYTLVRLPNTIYLF 272
Db 335 AIVLEWENKSGPDEL-QTYVQVYKPNANQAFSEISG-ITMTIYVRLSPPTETEYF 391
Qy 273 VRAINPKVSVTQXAPKNGSTWAVPLPPP--PVQPLPG-TELEHTAVQEQNGTSD- 328
Db 392 VIANV-----NIGRGPSAPATCTGEEKMESAPVQWVITLSSST 432
Qy 329 -----SNCPPLPQVTLHQGLDELEEDDORVPTPVKVG-----ASSPAISFGQ--- 374
Db 433 MYITWEP-----ETYPGCVIGYKIVYTTNSQPSASNSQWQDN 472
Qy 375 -----STALTP-----SPREMQPLMQASPYFT-----SSQRPRTSPF-----STDSVSAAL 419
Db 473 SELITVSDVTFAHITVYVQVATISWGAGPMSTPVQVQAGVPSQSNFPAITDGTAVT 532
Qy 420 SQSRSP 425
Db 533 LQMTKP 538

RESULT 5

NCM2_MOUSE

NCM2_MOUSE STANDARD; PRT; 837 AA.

035136; 035962;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE NEURAL CELL ADHESION MOLECULE 2 PRECURSOR (N-CAM 2) (R8-8 NEURAL CELL ADHESION MOLECULE) (R4812).

GN NCM2 OR OCAM OR NCAM.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RM [1]

RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).

RC STRAIN=BALB/C; TISSUE=OLFACTORY NEUROEPITHELIUM;

RX MEDLINE=97368238; PubMed=9221781;

RA Yoshihara Y., Kawasaki M., Tanada A., Fujita H., Hayashi H.,

RA Kagamiyama H., Mori K.;

RT *OCAM: A new member of the neural cell adhesion molecule family

RT related to zone-to-zone projection of olfactory and vomeronasal

RT axons.*;

RL J. Neurosci. 17:5830-5842 (1997).

RM [2]

RP SEQUENCE FROM N.A. (SHORT FORM).

RC STRAIN=C57BL/6J; TISSUE=OLFACTORY EPITHELIUM;

RX MEDLINE=97476194; PubMed=9334170;

RA Alenius M., Bohm S.;

RT *Identification of a novel neural cell adhesion molecule-related gene

RT with a potential role in selective axonal projection.*;

RL J. Biol. Chem. 272:26083-26086 (1997).

CC -1: FUNCTION: MAY PLAY IMPORTANT ROLES IN SELECTIVE FASCICULATION AND ZONE-TO-ZONE PROJECTION OF THE PRIMARY OLFACTORY AXONS.

CC -1: SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT FORM).

CC -1: TISSUE SPECIFICITY: EXPRESSED IN SUBSETS OF BOTH OLFACTORY AND VOMERONASAL NEURONS IN A ZONE-SPECIFIC MANNER.

CC -1: SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -1: SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC

CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: AF001287; AA869125.1; -.

DR EMBL: AF001286; AA869124.1; -.

DR EMBL: AF016619; AAC53375.1; -.

DR MGD: MGI:1095738; OCAM.

DR INTERPRO: IPR001777; -.

DR INTERPRO: IPR003006; -.

DR PFM: PFM0041; fn3; 2.

DR PFM: PFM0047; ig; 5.

KW Cell adhesion; Transmembrane; Glycoprotein; Repeat;

KW Immunoglobulin domain; Signal; GPI-anchor; Alternative splicing.

FT SIGNAL 1 19

FT CHAIN 20 837

FT DOMAIN 20 697

FT TRANSMEM 698 718

FT DOMAIN 719 837

FT DOMAIN 35 100

FT DOMAIN 129 193

FT DOMAIN 225 288

FT DOMAIN 315 387

FT DOMAIN 415 481

FT DOMAIN 482 581

FT DOMAIN 594 678

FT DISULFID 42 93

FT DISULFID 136 186

FT DISULFID 232 281

FT DISULFID 322 380

FT DISULFID 422 475

FT CARBOHYD 177 177

FT CARBOHYD 219 219

FT CARBOHYD 309 309

FT CARBOHYD 406 406

FT CARBOHYD 419 419

FT CARBOHYD 445 445

FT CARBOHYD 474 474

FT CARBOHYD 562 562

FT VARSPLIC 694 837

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Query Match 12.84; Score 292.5; DB 1; Length 837;

Best Local Similarity 21.74; Pred. No. 1.2e-10;

Matches 101; Conservative 74; Mismatches 139; Indels 151; Gaps 17;

Qy 5 QGRVTVPFCTKGNPQAVFVWQESQNLFPNQ-----QPPNSRCSVSPGDLITNIQR 61

Db 224 RGEEMTLCKAGSGDPTISWFRNG--KLEENKILKISGW-----ELVITNIN 273

Qy 62 SDAGTITQCALVAGSILAKAQLVDTVLDTRPPPIIQGPANQVLAVDGTALLCATG 121

Db 224 RGEEMTLCKAGSGDPTISWFRNG--KLEENKILKISGW-----ELVITNIN 273

PT	SIGNAL	<1	2	POTENTIAL.
PT	CBAIN	3	1377	NEOGENIN.
PT	DOMAIN	3	1074	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	1075	1095	POTENTIAL.
PT	DOMAIN	1096	1377	CYTOPLASMIC (POTENTIAL).
PT	DOMAIN	36	105	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	135	197	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	232	296	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	324	386	IG-LIKE C2-TYPE DOMAIN.
PT	DOMAIN	405	502	FIBRONECTIN TYPE-III.
PT	DOMAIN	505	598	FIBRONECTIN TYPE-III.
PT	DOMAIN	599	698	FIBRONECTIN TYPE-III.
PT	DOMAIN	704	798	FIBRONECTIN TYPE-III.
PT	DOMAIN	819	919	FIBRONECTIN TYPE-III.
PT	DOMAIN	920	1021	FIBRONECTIN TYPE-III.
PT	DOMAIN	1087	1090	POLY-VAL.
PT	CARBOHYD	42	42	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	185	179	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	285	295	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	439	439	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	458	458	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	608	608	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	684	684	N-LINKED (GLCNAC. .) (POTENTIAL).
PT	CARBOHYD	878	878	N-LINKED (GLCNAC. .) (POTENTIAL).
SO	SEQUENCE	1377	AA:150637 MW: 8514 PROBABLY 6390 CIRC64:	

Query Match 12.5%; Score 284.5; DB 1; Length 1377;
Best Local Similarity 26.3%; Pred. No. 6.5e-10;
Matches 91; Conservative 55; Mismatches 153; Indels 47; Gaps 11;

QY	6	GRITVFPCEKGNQKQPAVFWKQESQNLKIPNPOQPNKSRVCSVPDQITNORSDAG	65
Db	232	QGSAAVPCVAGSLPAPVFNWKN - NEDVL - - - - - DTSSGRLALAGGSLEISDVTEDAG	285
QY	66	TYIQCALVAGSLARAAQLEVDVTDTRPPNIIQLGQANLAVDGTALLAKKATGDPL	125
Db	286	TYECVANDMGNTIEAQAQLTV - - - - - QVPPFLPQGANITAREMSDIVEFCEVTKPAG	339
QY	126	VISNLKRGDTPFGCDPRATIQEQGTQIKRLISDGTQITCVATSSSGEASWASLVDT	185
Db	340	TYWVKNYGVDPISDPIFKVYKE - NLQVLGLVSKSDGSPQCIANDGVNAGAGALLIE	398
QY	186	SGATISKYNDLSDLPSPSKQVTDVTSNVTLSWQPGTGTLPASATIIIEAFSPQSVNS	245
Db	399	HAPAT - - - - - GLPSPAPROGVAISVSTRKILWR - TPASDPHG - - - - - DNLTVSVFT	447
QY	246	WQVVA - NRVHT - - - - - LTVLGRLRNTHILFWRAINPVSQVTKXQPLNASTWNY	298
Db	448	KDYARERFVNTSQDQEMVTQJNMPATVYFKWNAKNGSGESAPLAVETQPEVOL	507
QY	298	PLPPPPV - - - - - PLRGT - ELEYHVADEQDQENSD	328
Db	508	PCPAPNIBAYTSPTSTVYTWKPTLSGNGTQNYKLVYKSGDNDK	553

```

RESULT 7
NEOI_CHICK
ID NEOI_CHICK STANDARD; PRT: 1443 AA.
AC Q90610;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEIGHORN; TISSUE=EMBRYONIC BRAIN;
RX MEDLINE=95105243; PubMed=7806578;
RA Vlietinck J., Roman J.M., Drever W.J.:

```

Qy	6	GRVTFPFCYTKGNQPAVWQKESGNLFFMQPQPNRSCVSPSTDLTITNQASD	65
Db	229	GQNVAFFVAGGPTTPITVWTKNGEELI-----TEDESERFALRAGSLLSIDVTEYVG	282
Qy	66	TYTCALITVAGSILAKAELEVTIDVLTRPPPIILGQPNATLVDVTKALLCKNDGPI	125
Db	283	TYTICLADNENETIEQAELAV-----QVPEELRKPANITAHESNDIVFCEVTKPT	336
Oy	126	VISWLKEGFTFPGDRPATIROGTLQKNLISDGTTCVATVSSGASMSAVLDVTE	185

```

REUSE 8
NEOL_MOUSE
ID NEOL_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEOI OR NEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP Sequence from N.A. and ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=97407661; PubMed=9264410;
RT Keeling S.L., Gad J.K., Cooper H.M.;
RA "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RL expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC
CC -1- SUBCELLULAR-LOCATION: TYPE 1 INTEGRAL MEMBRANE PROTEIN.
CC
CC -1- ALTERNATIVE-PRODUCTS: AT LEAST 5 ISOFORMS: 1 (SHOWN HERE), 2, 3,
CC AND 5, ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC
CC -1- DEVELOPMENTAL STAGE: EXPRESSED OBVIOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.
CC AND E16.5.
CC
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC or send an email to license@isb-sib.ch).
CC

```

DR EMBL: Y0535; CAA70727.1; -.
 DR HSSP: P02751; 1TTG.
 DR MGD: MGI:1097159; NBD1.
 DR INTERPRO: IPR001777; -.
 DR INTERPRO: IPR003006; -.
 DR PFAM: PF00041; fn3; 6.
 DR PFAM: PF00047; ig; 4.
 DR PRINTS: PR00014; FMTYPEIII.
 KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 36 POTENTIAL.
 FT CHAIN 37 1493 NEOGENIN.
 FT DOMAIN 37 1136 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1137 1157 POTENTIAL.
 FT DOMAIN 1158 1493 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 78 147 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 177 239 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 274 338 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 366 428 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 467 564 FIBRONECTIN TYPE-III.
 FT DOMAIN 567 660 FIBRONECTIN TYPE-III.
 FT DOMAIN 661 760 FIBRONECTIN TYPE-III.
 FT DOMAIN 766 860 FIBRONECTIN TYPE-III.
 FT DOMAIN 861 961 FIBRONECTIN TYPE-III.
 FT DOMAIN 962 1083 FIBRONECTIN TYPE-III.
 FT DOMAIN 1149 1153 POLY-VAL.
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 442 461 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 863 878 MISSING (IN ISOFORM 3).
 FT VARSPPLIC 1086 1096 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 1279 1331 MISSING (IN ISOFORM 5).
 SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5817C0E CRC64;

Query Match 12.2%; Score 279; DB 1; Length 1493;
 Best Local Similarity 24.3%; Pred. No. 1.5e-09;
 Matches 112; Conservative 66; Mismatches 191; Indels 92; Gaps 18;

Qy 6 GRIVYPCETKGNQPAVFWKREGSNLLFPWQPGQNSKVSYPICDLITNIQASDAG 65
 : : | | | | | : : : : : : : : : : | |
 Db 274 QSAVILPCVSLGPAVVRMK-NEEV----DTSSGLVLAGGLCEISVDVEDG 327
 66 YIIQALIVAGSILAKAQLEVDITDRPPIILQIPQANQVLTGALKAKRAGPLP 125
 : : | : : | : | : | : | : | : | : : : : : : : : : : : : : :
 328 TYFC-IADGKNTVEAQE----LTVQVPGFLQKAPNIATHEKSDIVCEVDIKPTP 381
 Qy 126 VISHLKGPTPGPRORATIQDGLTIKNLRISDGTGTYCVATSSSGEASVAVLWTE 185
 : : | : | : : | : | : | : | : | : | : : : : : : : : : : : :
 Db 382 PVANXGNDVVPISDNKIVKEH-NLVGLVKSDSGFQCIAEDNDVGNAGAAQLILE 440
 Qy 186 SGATIKSNYDLDLPGPPKQPVDTKSNVLSWGTPGTPLASATIAIIR---FSQSV 242
 : : | : | : : | :
 Db 441 -----HDVALPLPPT-SLTSATDHLA-----PATGFLSAPRDVAGSLVSTFT 486
 Qy 243 SINSQTVAN--HWKTLT-----TYVGLRPNIIYLFWRA 275
 : : | : | : | : | :
 Db 487 KLNKRPASDFGHDNLITSVFYTKBGVDRERVENTSQPGEMQVITQNLMPATYIKFWA 546
 Qy 276 INPKVSYTQXKPKQNNNGSWANVLPPLPPVQ-----PLPGT-ELERYAVE 319
 : : | : | : : | :
 Db 547 QNKRGSGESSALRVETQEVQLGPAFNIRAYATSPSITVTWETPLMCEGQNTKLY 606
 Qy 320 QQBNQDSDSCNCPPLPVQTYLAQLEDEDDRDVPTPVRGVA---SSPAISFGQST 376
 : : | : : : : | : | :
 Db 607 YMEGTQKQDQI-DVSSHSYTLNGLKRYTEYS-----FRVAYNKHGPOVSTQDAV 657

Qy 377 ATLTPSPREEMQPM---LQASPKFTSQRPRTPSPFSTDSN 414
 : : | :
 Db 658 RTISDVTPSAAPQNLSEVANSRIVHWQP-----PSSTQTN 694

RESULT 9
 CAML_HUMAN
 ID CAML_HUMAN STANDARD; PRT; 1257 AA.
 AC P32004;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEUTRAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
 GN LICAM OR CAML OR MIC5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92031698; PubMed-1932117;
 RX Kobayashi M., Miura M., Asou H., Uyemura K.;
 RT "Molecular cloning of cell adhesion molecule L1 from human nervous
 RT tissue: a comparison of the primary sequences of L1 molecules of
 RT different origin.";
 RL Biochim. Biophys. Acta 1090:238-240(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rosenthal A., Coutelle O., Drescher B.;
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92329295; PubMed-1627459;
 RA Reid R.A., Hemperly J.J.;
 RT "Variants of human L1 cell adhesion molecule arise through alternate
 RT splicing of RNA.";
 RL J. Mol. Neurosci. 3:127-135(1992).
 RN [4]
 RP SEQUENCE OF 353-1176 FROM N.A.
 RX MEDLINE-92020233; PubMed-1923824;
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
 RT "PCR walking from microdissection clone M54 identifies three exons
 RT from the human gene for the neural cell adhesion molecule L1
 RT (CAM-L1).";
 RL Nucleic Acids Res. 19:5395-5401(1991).
 RN [5]
 RP SEQUENCE OF 332-371 FROM N.A.
 RX MEDLINE-90353957; PubMed-2387585;
 RA Djabbal M., Mattel M.-G., Nguyen C., Roux D., Demengeot J.,
 RA Deniot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
 RT "The gene encoding L1, a neural adhesion molecule of the
 RT Immunoglobulin family, is located on the X chromosome in mouse and
 RT man.";
 RL Genomics 7:587-593(1990).
 RN [6]
 RP SEQUENCE OF 1030-1257 FROM N.A.
 RX MEDLINE-91132183; PubMed-1993895;
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
 RA Stallcup W.B.;
 RT "Isolation and sequence of the partial cDNA clones of human L1: homology
 RT of human and rodent L1 in the cytoplasmic region.";
 RL J. Neurochem. 56:797-804(1991).
 RN [7]
 RP SEQUENCE OF 20-36.
 RX MEDLINE-88298876; PubMed-3136168;
 RA Wolff J.M., Frank R., Mirjo K., Spiro R.C., Reisfeld R.A.,
 RA Rathjen F.G.;
 RT "A human brain glycoprotein related to the mouse cell adhesion
 RT molecule L1.";
 RL J. Biol. Chem. 263:11943-11947(1988).
 RN [8]
 RP VARIANT HSAS TYP-264.
 RX MEDLINE-94004956; PubMed-8401576;
 RA Jouet M., Rosenthal A., Macfarlane J., Kenrick S., Donnai D.;

RT "A missense mutation confirms the LI defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANT HSAS/MASA LEU-1194.
RX MEDLINE=95187172; PubMed=7881431;
RA Fransen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the LICAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE=95004608; PubMed=7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metznerberg A., Ionescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPGL), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the LI gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE=95004609; PubMed=7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
RA Reyniers E., Korn B., Poustka A., Wilson G., Schrander-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT LICAM.";
RL Nat. Genet. 7:408-413(1994).
RN [12]
RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
RX MEDLINE=95282776; PubMed=7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Nadelius C., Kenwick S.;
RT "New domains of neural cell-adhesion molecule LI implicated in
RT X-linked hydrocephalus and MASA syndrome.";
RL Am. J. Hum. Genet. 56:1304-1314(1995).
RN [13]
RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
RX MEDLINE=96153146; PubMed=8556302;
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, LI.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]
RP VARIANTS HSAS/MASA/SPGL SER-179 AND ARG-370.
RX MEDLINE=96037511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Frys J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in LI-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the LICAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; Y-439; T-443 DEL; C-784 AND L-936 -L-948 DEL.
RX MEDLINE=97386664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Peyps M.E., Ramsden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.N., Berry C., Goudie D.,
RA Moncla A., Dunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel LI CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).

RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the LI cell
RT adhesion molecule (LICAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE=98112489; PubMed=9452110;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
RL Hum. Mutat. Suppl. 1:S284-S287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE=98415726; PubMed=9744477;
RA Saugier-Verber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jouveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosti M., Frebourg T.;
RT "Identification of novel LICAM mutations using fluorescence-assisted
RT mismatch analysis.";
RL Hum. Mutat. 12:259-266(1998).
CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO ANIONIN ON NEURONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC -1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF THREE X-LINKED
CC SYNDROMES. 1: HYDROCEPHALUS DUE TO STENOSIS OF THE AQUEDUCT OF
CC SYLVUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
CC Milder PRESENTATION AND A LONGER LIFE EXPECTANCY. 3: SPASTIC
CC PARAPLEGIA TYPE 1 (SPGL). COLLECTIVELY THESE SYNDROMES ARE ALSO
CC KNOWN AS CRASH SYNDROME, AN ACROMYIA WHICH STANDS FOR CORPUS
CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,
CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
CC -1- DISEASE: DEFECTS IN LICAM ARE THE CAUSE OF HIRSCHSPRUNG DISEASE
CC (HSCR).
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=LICAM; NOTE=LICAM mutation Web Page;
CC WWW="http://hgins.ula.ac.be/dna/lab/11".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR ENSL: X59847; CAA42508.1; ..
DR ENSL: Z29373; CAA82564.1; ..
DR ENSL: M74387; AAA59476.1; ..
DR ENSL: X58775; CAA41576.1; ..

Query Match 11.8%; Score 268; DB 1; Length 1257;
Best Local Similarity 23.8%; Pred. No. 5.7e-09;
Matches 97; Conservative 46; Mismatches 143; Indels 122; Gaps 12;

QY 66 GRTVTPCCAGNQPQAVFWKRGSSQLNFPQPPQPNRSCVSPGDTLITINIQSDAG 65
Db 347 GTARLDQWQGRPPQETVIRWING----IPEELAKDQTRIQI-RGALLISNVQPSDTM 400

QY 66 YTIQALIVAGSILAKAQLEVIDLIDPPPIILQGPANQTLAVDG-TALLKCKATGDIPL 124
Db 401 VTQCEARNRREILLANVITVYQW----PAKILTADNQTHAVQGSTATLLCKAGPAPY 455

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

DR EMBL: U61262; AAB17263.1; -.
DR EMBL: U72391; AAC51287.1; -.
DR MIM: 601907; -.
DR HSSP: P02751; 1TTG.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 6.
DR PRFAM: PF00047; Ig; 4.
DR PRFAM: PF00014; FN1PEH11.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1461 NEOGENIN.
FT DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1106 1126 POTENTIAL.
FT DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 67 136 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 166 228 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 263 327 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 355 417 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 436 533 FIBRONECTIN TYPE-III.
FT DOMAIN 536 629 FIBRONECTIN TYPE-III.
FT DOMAIN 630 729 FIBRONECTIN TYPE-III.
FT DOMAIN 735 829 FIBRONECTIN TYPE-III.
FT DOMAIN 850 950 FIBRONECTIN TYPE-III.
FT DOMAIN 951 1052 FIBRONECTIN TYPE-III.
FT DOMAIN 1118 1121 POLY-VAL.
FT DISULFID 74 129 BY SIMILARITY.
FT DISULFID 173 221 BY SIMILARITY.
FT DISULFID 270 320 BY SIMILARITY.
FT DISULFID 362 410 BY SIMILARITY.
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1248 1300 MISSING (IN ISOPFORM 2).
FT CONFLICT 168 168 G -> N (IN REF. 2).
SO SEQUENCE 7461 AA: 159958 MW: 7AA8897E69635a21 CRC64:

```

Query Match 11.7%; Score 267.5; DB 1; Length 1461;
Best Local Similarity 24.6%; Pred. No. 7.4e-09;
Matches 86; Conservative 50; Mismatches 160; Indels 53; Gaps 9;

[illegible]


```

RESULT 12
CAML_RAT
ID CAML_RAT STANDARD; PRT; 1259 AA.
AC Q05695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1

```

RL Oncogene 11:2243-2254(1995).

```

RESULT 14
NRG_DROME
ID NRG_DROME STANDARD; PRT; 1239 AA.
AC P20241; Q24414;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGLIAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (Fruit fly).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Schizophora; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RT SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
EX MEDLINE=90030418; PubMed=2805067;
RA Bleher A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglial: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT L1."
RL Cell 59:447-460(1989).
RN [2]
RT SEQUENCE OF 1182-1239 FROM N.A.
EX MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bleher A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglial."
RL Neuron 4:697-709(1990).
RN [3]
RT X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
EX MEDLINE=94213741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bleher A.J., Bjorkman P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglial at 2.0 A."
RL Neuron 12:717-731(1994).
CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL

```

RESULT 15
 AXOL_HUMAN
 ID AXOL_HUMAN STANDARD; PRT; 1040 AA.
 AC Q02246;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1993 (Rel. 38, Last annotation update)
 DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL
 DE GLYCOPROTEIN 1).
 GN TAXI OR TAGI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Dutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE-93145965; PubMed-8425542;
 RA Haeleir T.H., Rader C., Stoekli E.T., Zuellig R.A., Sonderegger P.;
 RT "cDNA cloning, structural features, and eucaryotic expression of
 RT human TAG-1/axonin-1".
 RL Eur. J. Biochem. 211:329-339(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE-94140394; PubMed-8307567;
 RA Tsiotra C.P., Karagozou D., Theodorakis K., Michaelidis M.T.,
 RA Modi W.S., Furley J.A., Jessel M.T., Papanthasakis J.;
 RT "Isolation of the cDNA and chromosomal localization of the gene
 RT (TAG1) encoding the human axonal glycoprotein TAG-1".
 RL Genomics 18:562-567(1993).
 CC -- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
 CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
 CC -- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
 CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
 CC -- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -- SIMILARITY: CONTAINS 4 TUBULININ TYPE III-LIKE DOMAINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X68274; CAA48335.1; -.
DR EMBL: X67734; CAA47963.1; -.
DR PIR: S28830; S28830.
DR MIM: 190197; -.
DR INTERPRO: IPR001777; -.

```

Job time: 1292 sec

Query Match 11.2%; Score 256; DB 1; Length 1040;
Best Local Similarity 24.4%; Pred. No. 2.4e-08;
Matches 108; Conservative 59; Mismatches 151; Indels 124; Gaps 22;

Search completed: January 22, 2001, 12:29:51

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:54:03 ; Search time 559.88 Seconds
(without alignments)
90.856 Million cell updates/sec

Title: US-09-540-245a-19
Perfect score: 2280
Sequence: 1 QIQAQGRVITVPCTCKGNPQ.....TSAALSQSRPPTKKHKGG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Aligned: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTRMBL15:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organella:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1479	64.9	285	4	Q43608	Q43608 homo sapien
2	1409	61.8	1060	11	Q9Q2I3	Q9Q2I3 rattus norv
3	913	40.0	1651	4	Q9V6N7	Q9V6N7 homo sapien
4	911	40.0	1612	11	O89026	O89026 mus musculu
5	911	40.0	1651	11	O55005	O55005 rattus norv
6	755.5	33.1	1344	11	Q9Z2I4	Q9Z2I4 mus musculu
7	579	25.4	1273	5	O44928	O44928 caenorhabdi
8	545	23.9	1395	5	O44924	O44924 drosophila
9	545	23.9	1395	5	Q9W2I3	Q9W2I3 drosophila
10	478	21.0	823	5	Q9VQ10	Q9VQ10 drosophila
11	417	18.3	859	5	Q9V826	Q9V826 drosophila
12	388	17.0	874	5	O01632	O01632 caenorhabdi
13	336.5	14.8	2221	5	Q9U1M1	Q9U1M1 drosophila
14	331.5	14.5	2222	5	O87394	O87394 drosophila
15	330	14.5	1788	13	Q9IAJ0	Q9IAJ0 xenopus lae
16	330	14.5	1904	11	O64699	O64699 mus musculu
17	325	14.3	1499	13	Q90815	Q90815 gallus gall
18	323.5	14.2	1501	11	Q9QW00	Q9QW00 rattus norv
19	323.5	14.2	1863	11	Q64605	Q64605 rattus norv

20	322.5	14.1	1948	4	Q13332	Q13332 homo sapien
21	319.5	14.0	475	4	O75255	O75255 homo sapien
22	316	13.9	1894	11	Q64487	Q64487 mus musculu
23	314	13.8	1502	4	Q9UM81	Q9UM81 homo sapien
24	308.5	13.5	1896	13	Q9IAJ1	Q9IAJ1 xenopus lae
25	305.5	13.4	1898	11	Q64604	Q64604 r protein-t
26	300.5	13.2	1598	4	Q9P2I4	Q9P2I4 homo sapien
27	297.5	13.0	1232	13	Q90284	Q90284 carassius a
28	297	13.0	1264	5	P91767	P91767 manduca sex
29	297	13.0	1887	11	Q9QW67	Q9QW67 rattus sp.
30	296.5	13.0	1026	11	Q62845	Q62845 rattus norv
31	296.5	13.0	1277	13	Q98902	Q98902 fugu rubrip
32	288	12.6	2016	5	Q9V4J9	Q9V4J9 drosophila
33	288	12.6	2016	5	Q9NB41	Q9NB41 drosophila
34	287.5	12.6	1180	4	O15051	O15051 homo sapien
35	285.5	12.5	1299	4	Q92823	Q92823 homo sapien
36	285.5	12.5	1496	4	Q92626	Q92626 homo sapien
37	284.5	12.5	1377	11	P97603	P97603 rattus norv
38	283.5	12.4	1099	11	P97527	P97527 rattus norv
39	282	12.4	1443	13	Q90610	Q90610 gallus gall
40	281.5	12.3	1236	4	Q9U8I3	Q9U8I3 homo sapien
41	281.5	12.3	1308	4	Q9U8I4	Q9U8I4 homo sapien
42	281	12.3	793	11	O70246	O70246 mus musculu
43	279	12.2	1493	11	P97798	P97798 mus musculu
44	278	12.2	920	4	Q9P232	Q9P232 homo sapien
45	276	12.1	1028	11	Q62682	Q62682 rattus norv

ALIGNMENTS

RESULT 1
Q43608
ID Q43608 PRELIMINARY; PRT; 285 AA.
AC Q43608;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ROUNDABOUT 2 (FRAGMENT).
GN ROUNDABOUT 2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR ENSL: AF040991; AAC39576.1; .
DR INTERPRO: IPRO01777; .
DR INTERPRO: IPRO03006; .
DR PFAM: PF00041; fn3; 1.
DR PFAM: PF00047; lg; 2.
FT NON_TER 1 1
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 30606 MW; 05DF916A3DBA96C CRC64;

Query Match 64.9%; Score 1479; DB 4; Length 285;
Best Local Similarity 100.0%; Pred. No. 8.9e-104;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIQAQGRVITVPCTCKGNPQPAVFKWQSGNQLLPNQPOQNSRCSVSPITGLITNIQ 60
Db 1 QIQAQGRVITVPCTCKGNPQPAVFKWQSGNQLLPNQPOQNSRCSVSPITGLITNIQ 60
Qy 61 RSDAGYITQALIVAGSILAAQLEVTDLTRPPPIILQGPANQTLAVGSLKCAKAT 120
Db 61 RSDAGYITQALIVAGSILAAQLEVTDLTRPPPIILQGPANQTLAVGSLKCAKAT 120

Qy 121 GDPLPVIISWLBKGFPTGPRDPRATIQEQGTQLQKNLRISDGTGYTCVATSSSGEASNSAV 180
 Db 121 GDPLPVIISWLBKGFPTGPRDPRATIQEQGTQLQKNLRISDGTGYTCVATSSSGEASNSAV 180
 Qy 181 LDVTEGATISKNYDLSLPGPPSKQVYDVTKNSTVLSWQPGTGLPASAYIIIEAFS 240
 Db 181 LDVTEGATISKNYDLSLPGPPSKQVYDVTKNSTVLSWQPGTGLPASAYIIIEAFS 240
 Qy 241 SVNSNQTVANHVKTLLTYVGLRPNTIYLFWRINPKVSVTQ 284
 Db 241 SVNSNQTVANHVKTLLTYVGLRPNTIYLFWRINPKVSVTQ 284

RESULT 2

O9QZ13

ID O9QZ13 PRELIMINARY; PRT; 1060 AA.

AC O9QZ13;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

REPUISIVE GUIDANCE RECEPTOR (FRAGMENT).

Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Duteleostomi;

RA Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99200391; PubMed=10102268;

RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,

RA Tessier-Lavigne M., Kidd T.;

RT "Slit proteins bind Robo receptors and have an evolutionarily

conserved role in repulsive axon guidance.";

RL Cell 96:795-806(1999).

DR EMBL: AF182037; AAF04558.1; .

DR HSSP: P56276; IILK.

DR INTERPRO: IPR001547; .

DR INTERPRO: IPR001777; .

DR INTERPRO: IPR003006; .

DR PFAM: PF00041; fn3; 3.

DR PFAM: PF00047; Ig; 5.

DR PRINTS: PR00014; FNTYPEIIL.

DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNWN_1.

KW Receptor.

FT NON_TER 1060 1060

SQ SEQUENCE 1060 AA; 116790 MW; C4BC9C11E8542DAA CRC64;

Query Match 61.8%; Score 1409; DB 11; Length 1060;
 Best Local Similarity 64.7%; Pred. No. 8.1e-98;
 Matches 303; Conservative 28; Mismatches 81; Indels 56; Gaps 10;

Qy 1 QIQAQGRVTFPCETKGNPQPAVFWKQESGNLLFPNQPPNSRCVSYPGDLITINQ 60
 Db 326 QIQAQGRVTFPCETKGNPQPAVFWKQESGNLLFPNQPPNSRCVSYPGDLITINQ 385
 Qy 61 RSDAGTYICQALTAVGSIKAKAQLVTVLDRPPPIIQGPANQTLAVDGTALLKCAK 120
 Db 386 RSDAGTYICQALTAVGSIKAKAQLVTVLDRPPPIIQGPINQTLAVDGTALLKCAK 445
 Qy 121 GDPLPVIISWLBKGFPTGPRDPRATIQEQGTQLQKNLRISDGTGYTCVATSSSGEASNSAV 180
 Db 446 G-PLPVIISWLBKGFPTGPRDPRATIQEQGTQLQKNLRISDGTGYTCVATSSSGEASNSAV 504
 Qy 181 LDVTEGATISKNYDLSLPGPPSKQVYDVTKNSTVLSWQPGTGLPASAYIIIEAFS 240
 Db 505 LDVTEGATISKNYDLSLPGPPSKQVYDVTKNSTVLSWQPGTGLPASAYIIIEAFS 564
 Qy 241 SVNSNQTVANHVKTLLTYVGLRPNTIYLFWRINPKVSVTQXKPKQKNGSTWNVNPL 300
 Db 565 SVNSNQTVANHVKTLLTYVGLRPNTIYLFWRINPKVSVTQXKPKQKNGSTWNVNPL 620
 Qy 301 PPP-----PVQPLPGTELEHVAEQBNGDSWCCPPLVQT 338

Db 621 SPQAQGVDRQVQKELGDTVRLHNPVLPITVQVTVDRQ-----PQPIQG 669
 Qy 339 Y-----LRLGLDEL--EEDDDRVPTTPVRGVASSPAISFGQOSTATLTPSPR-EMQPM 390
 Db 670 YRVMTQTSGLQASTVWQNLDAKVPTE-----RSANVLNLAGVYTIKVRPTNRFQGM 724
 Qy 391 LQASPTXTSSQ-----PRPTSPFTDSNTSAAQSQRPRPTKXHG 433
 Db 725 DSEKRTIRITTEAPASAPQSVTVLVGSHNSTSVISWVQPPADHQMG 772

RESULT 3

O9Y5N7

ID O9Y5N7 PRELIMINARY; PRT; 1651 AA.

AC O9Y5N7;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ROUNDABOUT 1.

GN ROBO1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Duteleostomi;

RA Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98117249; PubMed=9458045;

RA Kidd T., Brose K., Mitchell K.J., Petter R.D., Tessier-Lavigne M.,

RA Goodman C.S., Tessier G.;

RT "Roundabout controls axon crossing of the CNS midline and defines a

novel subfamily of evolutionarily conserved guidance receptors.";

RL Cell 92:205-215(1998).

DR EMBL: AF040990; AAC39575.1; .

DR HSSP: P56276; IILK.

DR INTERPRO: IPR001777; .

DR INTERPRO: IPR003006; .

DR PFAM: PF00041; fn3; 3.

DR PFAM: PF00047; Ig; 5.

SQ SEQUENCE 1651 AA; 180928 MW; 9D98DC7CAB73074D CRC64;

Query Match 40.0%; Score 913; DB 4; Length 1651;
 Best Local Similarity 23.2%; Pred. No. 2.8e-60;
 Matches 258; Conservative 54; Mismatches 118; Indels 684; Gaps 12;

Qy 1 QIQAQGRVTFPCETKGNPQPAVFWKQESGNLLFPNQPPNSRCVSYPGDLITINQ 60
 Db 360 QVVALGRTVTFPCETKGNPQPAVFWKQESGNLLFPNQPPNSRCVSYPGDLITINQ 419
 Qy 61 RSDAGTYICQALTAVGSIKAKAQLVTVLDRPPPIIQGPANQTLAVDGTALLKCAK 120
 Db 420 RSDAGTYICQALTAVGSIKAKAQLVTVLDRPPPIIQGPANQTLAVDGTALLKCAK 479
 Qy 121 GDPLPVIISWLBKGFPTGPRDPRATIQEQGTQLQKNLRISDGTGYTCVATSSSGEASNSAV 180
 Db 480 GSPVPTILNKGQVLSTQDSRIKQLGNGVLIYKALYSADTGRTYCIASPTSGEATWSAY 539
 Qy 181 LDVTEGATIS--KNYDLSLPGPPSKQVYDVTKNSTVLSWQPGTGLPASAYIIIEAF 238
 Db 540 LEVQPGVGVQPPRTDKNLIPASPKPEVTVSRNTVLSWQPNLNGATPTSVIIIEAF 599
 Qy 239 SQSVNSNQTVANHVKTLLTYVGLRPNTIYLFWRIN----- 277
 Db 600 SHAGSSNQTVANHVKTETSAIKGLKNATYLVRAANAGISDPQSIDPVPKTVQVLP 659
 Qy 278 ----- 277
 Db 660 TSQGVDRQVQKELGDTVRLHNPVLPITVQVTVDRQ-----PK 279
 Qy 278 ----- 279
 Db 720 ESDNLVFEVRIPAKNSWVLPDLKGVNVEIKARPFNEQADSEIKFATLEAPASAPP 779

Qy 280 VSVTQK- 286
 Db 780 QGVTVSKNDGNTAILVSWQPPEDTQGNQVQRYKVKCLNETRTHNKTVDSGTSFSVI 839
 Qy 287 -----PO----- 288
 Db 840 PFLVPGIRTSVEVAASGTAGSGVKSSEPPQITQDAHGNGVPSVEDQVLSAQGISDVVKQPAF 899
 Qy 289 -----KNG- 292
 Db 900 IAGIGAACWILLWFSINLYRRKRNGLSTYAGIRKVPSTFTPTVTYQRGESAVSSG 959
 Qy 293 -----STW- 297
 Db 960 GRPELLNISEPAAQPLWADTWPNTGNHNDSCISCTAGNCSNLTITSRPACTIANY 1019
 Qy 298 ----- 297
 Db 1020 NNQLDNKQTNLMPESTVYGDVLSNKNEMKTFNSPLNKDGRFVNPSSGQPTPYATTQLI 1079
 Qy 298 ----- 297
 Db 1080 QSLNNNNNGSGDSEKHKPLGQKQEVAPVQYVIVEQNKLAQDTRANDTVPPTIPTN 1139
 Qy 298 -----VP-----LPPVQVQLPGLTE 313
 Db 1140 QSTQDNTGGSYNSDRGSSSTSGSGHKKGARTPKVQKQGNMADLLPPPAHPPHNS 1199
 Qy 314 EHYAVQENGQYSDSNCPPLPVQTYLHQLGDELD-EDDDRVPPTVPVGVASSP-AISF 371
 Db 1200 EETNIVDS- YQEMPCVPVPAWYQQ--DELEEEDEERGPTVPVGAASSPAWY 1255
 Qy 372 GQSTATLTPSPREEMQPLQASP- 395
 Db 1256 SHQSTATLTPSQBELQPLQDCPETHGMHQDPRRRQVSPVPPPPISPPTHTYGIS 1315
 Qy 396 ----- 395
 Db 1316 GFLNSDMQDPAPEEEDAEADVAKMTRRLLLGQLTPASVGSDESSVTSGMNGW 1375
 Qy 396 -----XTSS 400
 Db 1376 SASEEDNISRRGSSVSSDGSFFTDADPAQNVAAAYTAGLKVARQMDAAGRPHAS 1435
 Qy 401 GRPRPTSPSTDSMTSAAALSQSRPRPKKHKG 434
 Db 1436 QCPRTSPVSTDSNMSAAVMKTRPAKLKHQPG 1469

RESULT 4
 089026
 ID 089026 PRELIMINARY; PRT; 1612 AA.
 AC 089026;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE DUTTI PROTEIN.
 GN ROBO1 OR DUTTI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Wu M.C., Lowe N., Fordham R., Rabbitts P.;
 RT "The mouse homologue of human DUTTI/R-robl gene: protein sequence and
 RT chromosomal location."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y17793; CAA76850.1;
 DR HSP: P56276; IRLK.
 DR MGD: MGI:1274781; Robol.

DR INTERPRO: IPR001777;
 DR INTERPRO: IPR003006;
 DR PFAM: PF00041; fn3; 3.
 DR PFAM: PF00047; lg; 5.
 SQ SEQUENCE 1612 AA; 176406 MW; 5P2988C544796B4B CRC64;

Query Match 40.0%; Score 911; DB 11; Length 1612;
 Best Local Similarity 22.8%; Pred. No. 3.9e-60;
 Matches 254; Conservative 55; Mismatches 120; Indels 686; Gaps 10;

Qy 1 QIVAGRTVTPCETXGNQPAVFWQKESQNLFPNQPPQNSRCSVPTGDLITNIIQ 60
 Db 321 QVVALGRTVTPQCEATGNQPAIFWRGSSQNLFFSTQPPQSSSRFSVSTGDLITNIIQ 380
 Qy 61 RSDAGYITQCALTVAGSILAKAQLVETDLTRPPPIILQGPAMQTLAVDGTALLKCAT 120
 Db 381 RSDGYYITQCLNVAGSIITKAYLETVDIADRPPIVQSGPVMTAVDGTLLSCVAT 440
 Qy 121 GDPLPVISMLAGEFTFPGRDPRATIQRGVLOIKNLRIISDTGYTTCVATSSSGEASHSAV 180
 Db 441 GSPARTILMRDGLVWSTQDSRIKQLESGVQLIRKALDQDGHYCTASTPSSGEATWSAY 500
 Qy 181 LDVTSRATIS--KNTDLSLPGPSPKQVDTYKNSVTLWQPGTPTGLPASAYITAP 238
 Db 501 IEVQEPGVQVPPRPTDPLPSAKSPKPEVTDVSKNVTLSWQNLGMSATPTSTIITAP 560
 Qy 239 SQSVSNSQTVANVKTITLYTVGLRNPITLYLWRAIN----- 277
 Db 561 SHAGSSWQTAAEWKTETFAKGLKPATLYLFLVAAANAYSIDSPQISDPVKTDQVPP 620
 Qy 278 ----- 277
 Db 621 TSGVDHKGQVRELGNVWHLNPTILSSSSVEVHWTVQDSQYIQGYIKLTPRSAGSHG 680
 Qy 278 -----PK 279
 Db 681 ESEWLVFVTRPTKNSVPIDLAKGVNVEIKARFPNFQADSEIKFANTLEAPSAPP 740
 Qy 280 VSVT- 283
 Db 741 RSVTVSKNDGNTAILVWQPPEDTQGNQVQRYKVKCLNETRTHNKTVDSGTSFSVI 800
 Qy 284 ----- 283
 Db 801 PSUVPGIRTSVEVAASGTAGGQVKSSEPPQITQDSHGNSVPSVEDQVLSAQGISDVVKQPAF 860
 Qy 284 ----- 283
 Db 861 IAGIGAACWILLWFSINLYRRKRNGLSTYAGIRKVPSTFTPTVTYQRGESAVSSG 920
 Qy 284 ----- 283
 Db 921 GRPELLNISEPAAQPLWADTWPNTGNHNDSCINCTAGNCSNLTITSRPACTIANY 980
 Qy 284 ----- 283
 Db 981 NNQLDNKQTNLMPESTVYGDVLSNKNEMKTFNSPLNKDGRFVNPSSGQPTPYATTQLI 1040
 Qy 284 -----QXKPO----- 288
 Db 1041 QANLNNNNNGAGDSSEKHKPLGQKQEVAPVQYVIVEQNKLAQDTRANDTVPPTIPTN 1100
 Qy 289 -----KNGSTWVNPPLPPVQLPGLTE 312
 Db 1101 QSTQDNTGGSYNSDRGSSSTSGSGHKKGARTPKVQKQGNMADLLPPPAHPPHNS 1159
 Qy 313 EHYAVQENGQYSDSNCPPLPVQTYLHQLGDELD-EDDDRVPPTVPVGVASSP-AIS 370
 Db 1160 SEETNIVDS- YQEMPCVPVPAWYQQ--DELEEEDEERGPTVPVGAASSPAWY 1215
 Qy 371 GQSTATLTPSPREEMQPLQASP- 395
 Db 1436 SHQSTATLTPSQBELQPLQDCPETHGMHQDPRRRQVSPVPPPPISPPTHTYGIS 1315

```

RESULT      6
Q92Z14
ID   Q92Z14      PRELIMINARY;          PRT;   1344 AA.
AC   Q92Z14;
DT   01-MAY-1999 (TrEMBL: 10, Created)
DT   01-MAY-1999 (TrEMBL: 10, Last sequence update)
DT   01-OCT-2000 (TrEMBL: 15, Last annotation update)
DE   RIG-1 PROTEIN.
GN   RIG1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDJB databases.
DR   EMBL; AF060570; AAD11628.1; -.
DR   HSSP; P56276; 1TLX.

```


Db 419 CSAFSDVSSVYVFLQVSSV-DERPPIIQIGPANQTLPGKSVATLPCATNGSPRIK 477

Qy 129 WLAEFGTTPGPRATIQDQGLQIKNLRIISDTGYTVAISSGSEASVAVLWTESSA 188

Db 478 WFDHGAQVAGN-RYSIIQSSSLVDDQLSDSGTTCYACGERGETSWAATLVKPKGS 536

Qy 189 T-ISKNYDLSLPGPSPKQVDTVKNSTVLS-----QPGTQPLASANIITAFSSGV 242

Db 537 TSLHRADPSTYAPPAGTPEKVLNVSRTSISLRWAKSQEPKGVG--PIIGTYEYFSD 594

Qy 243 SWSQWIVANHYKTYTLTVRGLPNTYLFVRAINPK-VSV 282

Db 595 QTGVIVAAHVGDTQVTISGLPGTSTVFLVRAINQGISV 635

RESULT 9

Q9W213

AC Q9W213 PRELIMINARY; PRT; 1395 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ROBO PROTEIN.

GN ROBO.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin J.K., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.N., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegyan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., Mpherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mosherif A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Stange R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Teh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AB003458; AAF46887.1; .

DR HSP: P56276; 17LK.
 DR FLYBASE: FBgn0005631; robo.
 DR INTERPRO: IPR001777; .
 DR INTERPRO: IPR003006; .
 DR PFM: PFM0041; fn3; 3.
 DR PFM: PFM0047; lg; 5.
 DR PRINTS: PR00014; FNTYPR11.
 SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13FO CRC64;

Query Match 23.9%; Score 545; DB 5; Length 1395;
 Best Local Similarity 40.9%; Pred. No. 1.1e-32;
 Matches 115; Conservative 47; Mismatches 105; Indels 14; Gaps 7;

Qy 9 VIFPCEKGNQPGVFWKESQNLFPQFQNGRCVSPGCDLIINIQROSDAGTY 68

Db 362 VQLPCWASNGPPFVYVFLQVSS-LDERPPIIQIGPANQTLPGKSVATLPCATNGSPRIK 477

Qy 69 QALTVAGSIIAKKQAVLDTLDRPPPIIQIGPANQTLVAGTALLKCAKGLDPLVPS 128

Db 419 CSAFSDVSSVYVFLQVSS-LDERPPIIQIGPANQTLPGKSVATLPCATNGSPRIK 477

Qy 129 WLAEFGTTPGPRATIQDQGLQIKNLRIISDTGYTVAISSGSEASVAVLWTESSA 188

Db 478 WFDHGAQVAGN-RYSIIQSSSLVDDQLSDSGTTCYACGERGETSWAATLVKPKGS 536

Qy 189 T-ISKNYDLSLPGPSPKQVDTVKNSTVLS-----QPGTQPLASANIITAFSSGV 242

Db 537 TSLHRADPSTYAPPAGTPEKVLNVSRTSISLRWAKSQEPKGVG--PIIGTYEYFSD 594

Qy 243 SWSQWIVANHYKTYTLTVRGLPNTYLFVRAINPK-VSV 282

Db 595 QTGVIVAAHVGDTQVTISGLPGTSTVFLVRAINQGISV 635

RESULT 10

Q9WQ10

AC Q9WQ10 PRELIMINARY; PRT; 823 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE C5G481 PROTEIN.

GN C5G481.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin J.K., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.N., Glasser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegyan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

RESULT 11
Q9VPZ6
ID Q9VPZ6 PRELIMINARY; PRT; 859 AA
AC Q9VPZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

QY 3 VAGGRTVTCEKTEKGNPQPAVWQKESQNLFLFNWQPPQMSRCSVPSGTD----LITIN 58
Db 269 VELGADTSPFCRAIENKPTFIWTKNNLSLIPGAD--PLDRFSLNTEEGHSILTR 326

QY 59 IQRSADAGTYI¹COALTVAGSILAKALEVTDVLTRPPPIILQGPANITLAVDGTALLKC 117
Db 327 FORTDKDLVLCIKNNEVAYSITSYVGLSI--DSOEDEPPIPIISGPNVITLPIKSLATLC 385

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang L.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE OF 1704-2221 FROM N.A.
 RC STRAIN=BERKELEY;
 RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
 RA Agbayani A., Arcaina T.T., Baxter E., Blazek R.G., Butenhoff C.,
 RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
 RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
 RA Houston K.A., Humasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
 RA Park S., Sequeira A., Sethi H., Sair E., Svirskas R.R., Weinburg T.,
 RA Celisner S.E.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132792; CAB65848.1;
 DR EMBL: AB003418; AAF45541.1;
 DR EMBL: AF132195; AAD34783.1;
 DR HSP: P56276; ITLX.
 DR FLYBASE: FBgn0021764; sdk.
 DR INTERPRO: IP0001777;
 DR INTERPRO: IP0003006;
 DR PFAM: PF00041; fn3; 13.
 DR PFAM: PF00047; ig; 5.
 DR PRINTS: PR00014; FWNPTEIII.
 KW Polymorphism.
 FT VARIANT 51 51 Y -> CYAD (IN STRAIN OREGON-R).
 SQ SEQUENCE 2221 AA; 245963 MW; 67451AD6A57F0660 CRC64;

Query Match 14.8%; Score 336.5; DB 5; Length 2221;
 Best Local Similarity 38.0%; Pred. No. 9.2e+17;
 Matches 109; Conservative 56; Mismatches 141; Indels 83; Gaps 16;

QY 6 GRVTFPCETKGNQPAVFWQKBSQNLFPNQPNRSRCSVPTGLITITNQSDAG 65
 DQ 466 GDAATSCRAVGSFNPNTW-----TNYTQVDSISRWQLBSGDLISINISVDAP 518
 QY 66 YIIQALIVAGSILAKAQLEVDVLTDRPPIILQGFANQVLAVDGTALKKATGDP-L 124
 DQ 519 LITCVANRAGSGVAGYLSVL-VRTQ-----TIQPVDTIYLLGLATLQCKVSSDPV 572
 QY 125 PV-ISWLKSG---PFFPGDPRATIQGTLQIKNLISDTGTTTCVATSSSGASNAVL 181
 DQ 573 PTINDWREGQSTPISNSQRIGVQADQLIQAVRASDVGSYACVTPSPGNETRAARL 632
 QY 182 DVTESGATISKNTDLSLPGPPSKPVQ---TDVTKNSVLSWQSPGTPLASATII--- 235
 DQ 633 SVIE-----LFFPSSNVKVERLPEPQASINWSVTPGDSNPSISKFIQR 679
 QY 236 ----EAFSQSVSN---SQIVANHVKT---TLITVGLRPNITLFWNRAINPVQIX 285
 DQ 680 EVSELEKIVGVPDPLLNITELSNVSDQRWILLENKAATVYQFVSANV---RVGEG 736
 QY 286 KPNKNGSTWNVPL-----PP-----PPVOPLPGLTEHYAVEQGE 322
 DQ 737 SP--SEPSNVVLPQAPSGPPVGFVGSARMSSEITQWQPLLEHNSQLIGLITRLR 794
 QY 323 NCYDSDSCWCPPLPVQIYLLHQGLEDELED 351
 DQ 795 FGIINNVFMS-----YQNTINEAQRN 814

RESULT 14
 Q97394
 ID Q97394 PRELIMINARY; PRT: 2222 AA.

AC Q97394;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SIDEKICK PROTEIN.
 GN SDK.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nguyen D.M.T., Liu Y., Litsky M.L., Reinke R.;
 RT "Sidekick, a member of the immunoglobulin superfamily, is required for
 pattern formation in the *Drosophila* eye.";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U88578; AAD09632.1;
 DR HSP: P56276; ITLX.
 DR FLYBASE: FBgn0021764; sdk.
 DR INTERPRO: IP0001777;
 DR INTERPRO: IP0003006;
 DR PFAM: PF00041; fn3; 13.
 DR PFAM: PF00047; ig; 5.
 DR PRINTS: PR00014; FWNPTEIII.
 SQ SEQUENCE 2222 AA; 246174 MW; 18853CCAF98D3BC2 CRC64;

Query Match 14.5%; Score 331.5; DB 5; Length 2222;
 Best Local Similarity 27.3%; Pred. No. 2.2e+16;
 Matches 107; Conservative 58; Mismatches 138; Indels 89; Gaps 15;

QY 6 GRVTFPCETKGNQPAVFWQKBSQNLFPNQPNRSRCSVPTGLITITNQSDAG 65
 DQ 469 GDAATSCRAVGSFNPNTW-----TNYTQVDSISRWQLBSGDLISINISVDAP 521
 QY 66 YIIQALIVAGSILAKAQLEVDVLTDRPPIILQGFANQVLAVDGTALKKATGDP-L 124
 DQ 522 LITCVANRAGSGVAGYLSVL-VRTQ-----TIQPVDTIYLLGLATLQCKVSSDPV 575
 QY 125 PV-ISWLKSG---PFFPGDPRATIQGTLQIKNLISDTGTTTCVATSSSGASNAVL 181
 DQ 576 PTINDWREGQSTPISNSQRIGVQADQLIQAVRASDVGSYACVTPSPGNETRAARL 635
 QY 182 DVTESGATISKNTDLSLPGPPSKPVQ---TDVTKNSVLSWQSPGTPLASATII--- 235
 DQ 636 SVIE-----LFFPSSNVKVERLPEPQASINWSVTPGDSNPSISKFIQR 682
 QY 236 ----EAFSQSVSN---SQIVANHVKT---TLITVGLRPNITLFWNRAIN--- 277
 DQ 683 EVSELEKIVGVPDPLLNITELSNVSDQRWILLENKAATVYQFVSANV---RVGEG 742
 QY 278 PKYSVTQKPKXNNGSTWNVPLP-----PPVOPLPGLTEHYAVE 319
 DQ 743 EPSNVVLPQAPSGPPVGFVGSARMSSEITQWQPLLEHNSQLIGLITRLR 794
 QY 320 QBNQYDSDSCWCPPLPVQIYLLHQGLEDELED 351
 DQ 795 FGIINNVFMS-----YQNTINEAQRN 817

RESULT 15
 Q9IAJ0
 ID Q9IAJ0 PRELIMINARY; PRT: 1788 AA.
 AC Q9IAJ0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE RECEPTOR PROTEIN TYROSINE PHOSPHATASE LAR.
 GN XPTP-LAR.
 OS *Xenopus laevis* (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;

OX NCBI_TaxID=8355;

RN

RP

RX MEDLINE=20193505; PubMed=10727868;

RA Johnson K.G., Holt C.E.;

RT "Expression of CRYP-alpha, LAR, PTP-delta, and PTP-rho in the

RT developing xenopus visual system.";

RL Mech. Dev. 92:291-294 (2000).

DR EMBL: AF197945: AAF43606.1: -

SQ SEQUENCE 1788 AA; 200270 MW; AB192549866D9067 CRC64;

Query Match 14.5%; Score 330; DB 13; Length 1788;
Best Local Similarity 29.8%; Pred. No. 2.2e-16;
Matches 108; Conservative 47; Mismatches 154; Indels 54; Gaps 15;

Oy 1 QIVAQGRITWTFPCETKGNPQPAVFWQKEGSQLFPNQPPQPNRSRCSVSPTGLDITINIQ 60
::|:|||||::|:
Db 141 KVVVKTRTATMLCAASGNPDPEITWFKD-----FLPVDTASSNGRIKOLRSGALQIENSE 195

61 RSDAGYYICQALTVAGS-ILAKAQLEWTDVLTRPPPIILQGPANQTLAVDGTALLKCKA 119
 ||||| :|: ||| | | | :|: | :|: |||
 Db 196 ESDQGYKTECVATNSAGTRYSAIPANLYVR---VRRVAPRFSIPPSNHEVMPGGSVNLTCVA 252

Qy 120 TGDPLPVISWLGKEGTFPPGRDPRAITQEQTGLQIKNLRISDTGTYYTCVATSSSGEASWSA 179
| : | : | : | : : : | : : | : | : ||||| | | |
Db 253 VGPAMPYVKWM-AGLEELTKEDEMPVGRRNG-LELTN--IKDSANYTCVAISSLGM|--FA 306

```
Qy      180 VLDVTESGATISKNYDLSDLPGPPSKPQVDTVKNSVTLSWQPGTPTGLPASAYIIIEAFS 239  
        | :|          : ||||   ||:| |||:| | | | |:|  
Db      307 VAQIT-----VKALPKPLDAMVTETATSVTLTWDSGNPD--PVSYIVIOYKP 353
```

Oy 240 QSVSNHQIVANHVKTTLYTVRGLRPNTIYLPVRAIN----PKVSTQXKPQNKGST 294
 :: :| | : || |::|| :| |:| :| :| :| :| :| :| :| :| :| :|
Db 354 KASESSQEV-DGVATTRYISIGGLSPPESEYFRIIAVNINIGRGPPSEVIEAOTGEOAPS 412

Qy 295 WANVLP PPPVPQ---LPGTEHYAYEQQENG-----YDSWSGCPPLPVQTYLHQG 343
 || || | : : : || | : | || :
 Db 413 -----PPLKVOARMLSASTMLVQWDLPEEPNGOIRGFRVYYITD---PHLPFSMNOKH 463

Qy 344 LED 346
: |
Db 464 VVD 466

Search completed: January 22, 2001, 12:54:12
Job time: 2053 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:19:47 ; Search time 233.01 Seconds
(without alignments)
21.719 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAATAAAEYAGLKVARQK.....REALGQGVLDLRTPSDPR 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Aligned: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Match	Length	DB	ID	Description
1	761	100.0	148	20	Y13568	Mouse Robo 1 polyp
2	761	100.0	148	20	Y08406	Mouse partial ROBO
3	691	90.8	1649	20	Y08404	Human ROBO1 protei
4	691	90.8	1651	20	Y13566	Human Robo 1 polyp
5	108	14.2	434	20	Y13567	Human Robo 2 polyp
6	108	14.2	434	20	Y08405	Human partial ROBO
7	101	13.3	515	19	W72076	HSV-2 strain SBS C
8	99.5	13.1	308	13	R22248	Sequence of rye-gr
9	97.5	12.8	308	20	Y25601	Lolium sp. allerg
10	97.5	12.8	414	19	W72159	HSV-2 strain SBS C
11	97.5	12.8	414	19	W72139	HSV-2 strain SBS C
12	88.5	11.6	821	14	R35451	Mouse eps8. Mus m

13	88	11.6	980	19	W69741	SAPAP2 protein. H
14	88	11.6	980	19	W69743	SAPAP1 protein. H
15	85	11.2	2424	21	Y78901	Calcium channel al
16	84	11.0	541	19	W37148	Mammalian Ena (Men
17	83	10.9	1319	14	R47043	Mammalian son of s
18	83	10.9	1336	16	R84638	mSOS1 protein. Mu
19	82	10.8	903	19	W77289	Zebrafish differen
20	81.5	10.7	449	19	W59866	Amino acid sequenc
21	81.5	10.7	456	19	W59865	Amino acid sequenc
22	81	10.6	1105	19	W44864	Human TPC2 telomer
23	81	10.6	1105	20	W73958	Human TPC2 protein
24	80.5	10.6	197	21	Y75526	Neisseria meningi
25	80.5	10.6	224	20	Y35778	Chlamydia pneumoni
26	80.5	10.6	548	18	W07700	Human ETS2 repress
27	80	10.5	127	18	W31265	Neospora caninum a
28	80	10.5	314	21	Y32321	Soybean beta-carot
29	80	10.5	386	20	Y59881	Human normal uteru
30	79.5	10.4	384	18	W23278	Bordetella pertuss
31	79.5	10.4	493	20	Y05396	Human TIE ligand N
32	79.5	10.4	493	21	Y70745	PSDQ-3 protein enc
33	79.5	10.4	493	21	Y54999	Human scarface 1 p
34	79	10.4	118	18	W19831	Plasmid pSP2alpha
35	79	10.4	256	20	Y00923	Alternatively spli
36	79	10.4	276	20	Y00922	Human CLAR1 protei
37	79	10.4	609	14	R32131	CMV Colburn region
38	78.5	10.3	350	20	Y35922	Extended human sec
39	78.5	10.3	543	18	W07702	Mouse ETS2 repress
40	78.5	10.3	783	19	W37151	Mouse neural Mena+
41	78.5	10.3	787	19	W37152	Mouse neural Mena+
42	78.5	10.3	802	19	W37153	Mouse neural Mena+
43	78.5	10.3	1784	16	R77223	Tuberous sclerosis
44	78	10.2	1333	21	Y68820	Amino acid sequenc
45	77.5	10.2	390	20	Y35923	Extended human sec

ALIGNMENTS

RESULT 1
ID Y13568
AC Y13568; standard; Protein; 148 AA.
DT 30-JUL-1999 (first entry)
DE Mouse Robo 1 polypeptide.
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
modulation; nerve cell function.
XX
OS Mus sp.
XX
PN W09925833-AL.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PD 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338009/28.
DR N-PSDB; X55772.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 50; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active Robo (roundabout) on its surface. The method comprises modulating the effective amount of Comm polypeptide in contact with the cell, where the amount of expressed active Robo is specifically modulated inversely with the modulation of the effective amount of Comm in contact with the cell. The method is used to modulate the amount of active Robo expressed on a cell. The method can be used to screen for agents that modulate Robo:Comm interactions. This is particularly useful for modulating nerve cell function.

XX
SQ Sequence 148 AA;

Query Match 100.04; Score 761; DB 20; Length 148;
Best Local Similarity 100.04; Pred. No. 4.8e-71;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAIAAAAYAGLKVARRQMDAAGRHHFASQCPRTSPVSTDSNMSAVVIQKARPAKK 60
|||||
Db 1 aqavaaaayegilkvarrqmdaagrrhfahscprptspvstdsnmsavvikarpakk 60

61 QKHQPHLRREAYADLPPPPVPPPAIKSPTVQSKAQLEVRPMVKLASIEARTDRSSD 120
|||||
61 qkhqphlrrreyaddlppppvpppaiksptvqskaqlevrpmvklasiartdrssd 120

Qy 121 RKGSGYKREALDGRQVTLRTNPSDPR 148
|||||
Db 121 rkgsgygrealdgrqvtlrltnpsdpr 148

RESULT 2
Y08406
ID Y08406 standard; Protein; 148 AA.

XX
AC Y08406;
DT 24-JUL-1999 (first entry)
XX
DE Mouse partial ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX
OS Mus sp.
XX
PN W09920764-AL.
XX
PD 29-APR-1999.
XX
20-OCT-1998; 98WO-US22164.
14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.

PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;

DR WPI; 1999-312615/26.
DR N-PSDB; K57255.

PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 74-75; 80pp; English.

XX
CC This invention describes novel Robo (roundabout) polypeptides, involved in nerve guidance which have been isolated from *Drosophila* sp., *C. elegans*, human and murine samples. The products of the invention can be used to raise anti-Robo antibodies, which can be used to modulate cell function or morphology. The Robo polynucleotides and fragments are useful as probes and primers and for production of the Robo polypeptides. The probes and primers are also useful in screening assays.

SQ Sequence 148 AA;

Query Match 100.04; Score 761; DB 20; Length 148;
Best Local Similarity 100.04; Pred. No. 4.8e-71;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAIAAAAYAGLKVARRQMDAAGRHHFASQCPRTSPVSTDSNMSAVVIQKARPAKK 60
|||||
Db 1 aqavaaaayegilkvarrqmdaagrrhfahscprptspvstdsnmsavvikarpakk 60

Qy 61 QKHQPHLRREAYADLPPPPVPPPAIKSPTVQSKAQLEVRPMVKLASIEARTDRSSD 120
|||||
Db 61 qkhqphlrrreyaddlppppvpppaiksptvqskaqlevrpmvklasiartdrssd 120

Qy 121 RKGSGYKREALDGRQVTLRTNPSDPR 148
|||||
Db 121 rkgsgygrealdgrqvtlrltnpsdpr 148

RESULT 3
Y08404
ID Y08404 standard; Protein; 1649 AA.

XX
AC Y08404;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX
OS Homo sapiens.
XX
PN W09920764-AL.
XX
PD 29-APR-1999.
XX
20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X08404.

PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 65-71; 80pp; English.

XX
CC This invention describes novel Robo (roundabout) polypeptides, involved in nerve guidance which have been isolated from *Drosophila* sp., *C. elegans*, human and murine samples. The products of the invention can be used to raise anti-Robo antibodies, which can be used to modulate cell function or morphology. The Robo polynucleotides and fragments are useful as probes and primers and for production of the Robo polypeptides. The probes and primers are also useful in screening assays.

XX
SQ Sequence 1649 AA;

Query Match 90.84; Score 691; DB 20; Length 1649;
Best Local Similarity 88.54; Pred. No. 1.2e-62;
Matches 131; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 AQAIAAAAYAGLKVARRQMDAAGRHHFASQCPRTSPVSTDSNMSAVVIQKARPAKK 60
|||||
Db 1402 aqavaaaayegilkvarrqmdaagrrhfahscprptspvstdsnmsavvikarpakk 1461

Qy 61 QKHQPHLRREAYADLLPPPPVPPPAIKSPVQSKAQLVRFVWPKLASIEARTDRSSD 120
 Db 1462 Lkhqphlrrretyddlppppvpppaikspqtagktqlvrpvpvklpsmdartdrssd 1521

Qy 121 RKGSGYKRGALDGRQVTLRTNPSDPR 148
 Db 1522 rkqssykgrevldgrqvdmrtngpdr 1549

RESULT 4
 Y13566

ID Y13566 standard; Protein; 1651 AA.

XX
 AC Y13566;
 XX
 DT 30-JUL-1999 (first entry)

Human Robo 1 polypeptide.

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
 KW modulation; nerve cell function.

XX
 OS Homo sapiens.

XX
 PN W09925833-AL.

XX
 PD 27-MAY-1999.

XX
 PF 13-NOV-1998; 98WO-US24327.

XX
 PR 14-NOV-1997; 97US-0065543.

XX
 PA (REGC) UNIV CALIFORNIA.

XX
 PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX
 DR WPI; 1999-338008/28.
 DR N-PSDB; X55770.

XX
 PT Modulation of Robo-Comm polypeptide interactions

XX
 PS Disclosure; Page 44-48; 56pp; English.

XX
 CC The invention relates to a method for modulating the amount of Comm
 (commissureless) polypeptide in contact with a cell expressing active
 Robo (roundabout) on its surface. The method comprises modulating the
 effective amount of Comm polypeptide in contact with the cell, where the
 amount of expressed active Robo is specifically modulated inversely with
 the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 cell. The method can be used to screen for agents that modulate Robo:Comm
 interactions. This is particularly useful for modulating nerve cell
 function.

XX
 SQ Sequence 1651 AA;

Query Match 90.8%; Score 691; DB 20; Length 1651;
 Best Local Similarity 88.5%; Pred. No. 1.2e-62;
 Matches 131; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 AQAQAAAEYAGLKVARRQMDAGRHRFHASQCPRTSPVSTDSNMSAVVIQKARPAKK 60
 Db 1404 aqaqaaaeayaglkvarrqmdagrhrfhassqcprrtspvstdsnmsaavmqktrpakk 1463

Qy 61 QKHQPHLRREAYADLLPPPPVPPPAIKSPVQSKAQLVRFVWPKLASIEARTDRSSD 120
 Db 1464 Lkhqphlrrretyddlppppvpppaikspqtagktqlvrpvpvklpsmdartdrssd 1523

Qy 121 RKGSGYKRGALDGRQVTLRTNPSDPR 148
 Db 1524 rkqssykgrevldgrqvdmrtngpdr 1551

RESULT 5

Y13567

ID Y13567 standard; Protein; 434 AA.

XX

AC Y13567;

XX

DT 30-JUL-1999 (first entry)

XX

DE Human Robo 2 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
 KW modulation; nerve cell function.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

PT Misc-difference 285

FT /label= unknown

PT /note= "encoded by GTN"

PT Misc-difference 395

FT /label= unknown

PT /note= "encoded by NTT"

XX

PN W09925833-AL.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24327.

XX

PR 14-NOV-1997; 97US-0065543.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55771.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX

PS Disclosure; Page 49-50; 56pp; English.

XX

CC The invention relates to a method for modulating the amount of Comm
 (commissureless) polypeptide in contact with a cell expressing active
 Robo (roundabout) on its surface. The method comprises modulating the
 effective amount of Comm polypeptide in contact with the cell, where the
 amount of expressed active Robo is specifically modulated inversely with
 the modulation of the effective amount of Comm in contact with the cell.
 CC The method is used to modulate the amount of active Robo expressed on a
 cell. The method can be used to screen for agents that modulate Robo:Comm
 interactions. This is particularly useful for modulating nerve cell
 function.

XX
 SQ Sequence 434 AA;

Query Match 14.2%; Score 108; DB 20; Length 434;
 Best Local Similarity 57.9%; Pred. No. 0.0023;
 Matches 22; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 29 FHASQCPRTSPVSTDSNMSAVVIQKARPAKKQKHQP 66
 Db 397 ftssqcprrtspfdtdntsalsqsgprprtkhkgg 434

Y08405

RESULT 6

ID Y08405 standard; Protein; 434 AA.

XX

AC Y08405;

XX

DT 24-JUL-1999 (first entry)
 XX
 DE Human partial ROBO2 protein.
 XX
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
 KW cell morphology; screening assay.
 XX
 OS Homo sapiens.
 XX
 PN W09920764-A1.
 XX
 PD 29-APR-1999.
 XX
 PF 20-OCT-1998; 98WO-US22164.
 XX
 PR 14-NOV-1997; 97US-0971172.
 PR 20-OCT-1997; 97US-0062921.
 XX
 PA (REOC) UNIV CALIFORNIA.
 XX
 PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
 WPI: 1999-312615/26.
 DR N-PSDB; A51254.
 XX
 PT Robo polypeptides, a new immunoglobulin superfamily member
 XX
 PS Claim 1; Page 72-73; 80pp; English.
 XX
 CC This invention describes novel Robo (roundabout) polypeptides, involved
 CC in nerve guidance which have been isolated from *Drosophila* sp.,
 CC *C. elegans*, human and murine samples. The products of the invention can
 CC be used to raise anti-Robo antibodies, which can be used to modulate cell
 CC function or morphology. The Robo polynucleotides and fragments are useful
 CC as probes and primers and for production of the Robo polypeptides. The
 CC probes and primers are also useful in screening assays.
 XX
 SQ Sequence 434 AA;

Query Match 14.2%; Score 108; DB 20; Length 434;
 Best Local Similarity 57.9%; Pred. No. 0.0023;
 Matches 22; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 29 FHASQCPRPSPVSTDSNMSAVVIQAKAPAKKHQPG 66
 : : | | | | | | | | | | : | : | : | : |
 Db 397 ftsqrrprptpfstidntsaalsqsgrrprptkhhg 434

RESULT 7
 2075
 W72076 standard; Protein; 515 AA.

XX
 AC W72076;
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 96 ORF4 protein.
 XX
 KW HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor.
 XX
 OS Herpes simplex virus type 2.
 XX
 PN W09820016-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 31-OCT-1997; 97WO-US20016.
 XX
 PR 09-JUN-1997; 97US-0049018.
 PR 04-NOV-1996; 96US-0030279.
 XX

PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Chan JT, Dahrowski-Amarel CE, Delvecchio AM, Dillon SB;
 PI Esser KM, Leary JJ;
 XX
 DR WPI: 1998-286847/25.
 DR N-PSDB; V62150.
 XX
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal
 XX
 PS Claim 10; Page 70; 748pp; English.
 XX
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
 CC sequence of the invention. This sequence was isolated from a HSV-2 strain
 CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 96.
 CC Based on homology, this sequence is an immediate-early protein IE68.
 CC The proteins can be used for the treatment or prevention of disease, to
 CC induce an immunological response in a mammal or to identify inhibitors,
 CC activators or novel antivirals. Antagonists of the proteins can be used
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
 CC it can also be used to induce an immunological response in a mammal.
 XX
 SQ Sequence 515 AA;

Query Match 13.3%; Score 101; DB 19; Length 515;
 Best Local Similarity 27.7%; Pred. No. 0.014;
 Matches 31; Conservative 15; Mismatches 38; Indels 28; Gaps 4;

Qy 63 HQGHLRL-ENYADLDP-----PPVPPPAIKS---PTVQSKAQLEVRPVMW 105
 : : : : | : | | | | : | : | : | : | : | : |
 Db 90 hrcqmrqrgvatmadipdpavntpanhspgppppgskrrrrpvlpsseesekpote 149
 Qy 106 PKLAISEARTDRSDDRGSGYKGRALGRQVTLR-----TNPSD 146
 : : : : | : | : | : | : | : | : | : | : | : |
 Db 150 sssstssessedeagdlrgrrrrsprelgryrflsaeattgttesegtpsd 201

RESULT 8
 R22248
 ID R22248 standard; Protein; 308 AA.

XX
 AC R22248;
 XX
 DT 22-JUL-1992 (first entry)
 XX
 DE Sequence of rye-grass pollen allergens encoded by EcoRI insert
 DE lambda-12R.
 XX
 KW Rye grass pollinosis; diagnosis; therapy.
 XX
 OS Lolium perenne.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= signal
 XX
 PN W09203550-A.
 XX
 PD 05-MAR-1992.
 XX
 PF 16-AUG-1991; 91WO-A000369.
 XX
 PR 17-AUG-1990; 90AU-0001823.
 XX
 PA (TYME-) UNIV MELBOURNE.
 XX
 PI Singh MB, Hough T, Knox RB, Avjioglu A;
 XX
 DR WPI: 1992-096894/12.
 DR N-PSDB; Q23000.

XX
PT New nucleic acid sequences coding rye-grass pollen allergens -
PT esp. Lol p1a and Lol p1b and their fragments, for diagnosing and
PT detecting rye-grass pollinosis
XX
PS Disclosure; ; 81pp; English.
XX
CC The inventors claim a sequence encoding the rye grass pollen
CC allergen Lol p1a, or an antigenic fragment. The allergen can
CC alternatively be Lol p1b. The antigenic fragment has T-cell
CC stimulating activity and IgE stimulating activity. It does not bind
CC IgE specific for rye grass pollen however. It may be encoded by
CC clone 12R (Q23000) or 26.j (Q22246). R22248 is as printed in the
CC specification; several codons in Q23000 do not translate into the
CC amino acids written in the specification.
XX
SQ Sequence 308 AA;
Query Match 13.1%; Score 99.5; DB 13; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.011;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;

Qy 3 AVAAAEYAGLKVARRQMDAAGRHHFASQCPRPTSPVSTDSNMSAVVIQKAPAKKQ 62
| | | | | : : : | | | | | : | | | |
Db 39 atpaatpggwggedrraaaggrqlasrgpwppltp----- 78
Qy 63 HQGHRLREAYADLPPPPVPPAIPKSPVQSKAQLVRFVPMVKLASIEARTDSRRK 122
| | | : | | | | | | | : | | | | | : | |
Db 79 ----lrrtsststpspsppra-sptsaaka----pglipkl-----dtdayd-- 118
Qy 123 GGSYNGREALGRQVITLRTNP 144
: | | | | | | | |
Db 119 -vaykaeahprgqvrrlrhpc 139

RESULT 9
Y25601
ID Y25601 standard; protein; 308 AA.
XX
AC Y25601;
XX
DT 30-SEP-1999 (first entry)
XX
DE Lolium sp. allergen 2498581 Lol p 5a protein fragment.
XX
W Major histocompatibility complex; class II; desensitising; human;
W allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Lolium sp.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-GB00080.
XX
PR 21-SEP-1998; 98GB-0020474.
PR 09-JAN-1998; 98GB-0000445.
XX
(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Kay AB, Larche M;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens
XX
PS Example 6; Page 55; 117pp; English.

XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitizing patients to allergens present in e.g. grass, tree and weed
CC (including rawweed) pollens, fungi and moulds, foods, stinging insects,
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC Tenibrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents the Lolium sp. allergen 2498581 Lol p 5a.
XX
SQ Sequence 308 AA;

Query Match 12.8%; Score 97.5; DB 20; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.018;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;
Qy 3 AVAAAEYAGLKVARRQMDAAGRHHFASQCPRPTSPVSTDSNMSAVVIQKAPAKKQ 62
| | | | | : : : | | | | | : | | | |
Db 39 atpaatpggwggedrraaaggrqlasrgpwppltp----- 78
Qy 63 HQGHRLREAYADLPPPPVPPAIPKSPVQSKAQLVRFVPMVKLASIEARTDSRRK 122
| | | : | | | | | | | : | | | | | : | |
Db 79 ----lrrtsststpspsppra-sptsaaka----pglipkl-----dtdayd-- 118
Qy 123 GGSYNGREALGRQVITLRTNP 144
: | | | | | | | |
Db 119 -vaykaeahprgqvrrlrhpc 139

RESULT 10
W72159
ID W72159 standard; Protein; 414 AA.
XX
AC W72159;
XX
DT 08-JAN-1999 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 12 ORF41 protein.
XX
KW HSV-2 strain SB5; Immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
PN WO9820016-A1.
XX
PD 14-MAY-1998.
XX
PF 31-OCT-1997; 97WO-US20016.
XX
PR 09-JUN-1997; 97US-0043018.
PR 04-NOV-1996; 96US-0030279.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Chan JT, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JJ;
XX
DR WPI; 1998-286847/25.
DR N-PSDB: V62175.
XX
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal

Qv 62 KHOPGHLRREAYAD-DLPPP-----PVPPPAIKSPTVQSKA----- 96

DE
MM

KW Alpha 1A subunit; central nervous system calcium channel; dementia;
 KW Alzheimer's disease; calcium channel function evaluation;
 KW cerebral function.
 XX
 OS Xenopus sp.
 XX
 PN JP2000026315-A.
 XX
 PD 25-JAN-2000.
 XX
 PF 09-JUL-1998; 98JP-0194236.
 XX
 PR 09-JUL-1998; 98JP-0194236.
 XX
 PA (DAUC) DAIICHI PHARM CO LTD.
 XX
 DR WPI; 2000-176900/16.
 XX
 PT Evaluating calcium channel activator - involves measuring inhibition or
 PT change in functions of coupling between alpha subunit of G-protein and
 PT alpha-1 subunit of C-terminal area, used for treating dementia and
 PT improving cerebral functions
 XX
 PS Example 1; Page 9-10; 11pp; Japanese.
 XX
 CC This sequence represents the alpha 1A subunit of the Xenopus calcium
 CC channel of the central nervous system. The invention relates to the
 CC evaluation of a calcium channel activator. The evaluation process
 CC consists of measuring the inhibition or change in function of coupling
 CC between a G-protein alpha subunit and the alpha 1 subunit of the calcium
 CC channel of the central nervous system. Test compounds which are found to
 CC cause the inhibition of coupling or a change in function can be
 CC identified from tests carried out on genetically engineered cells. The
 CC evaluation method may be used to discover compounds for treating dementia
 CC and improving cerebral function in diseases such as Alzheimer's disease.
 XX
 SQ Sequence 2424 AA;

Query Match 11.2%; Score 85; DB 21; Length 2424;
 Best Local Similarity 25.6%; Pred. No. 3.9;
 Matches 34; Conservative 15; Mismatches 46; Indels 38; Gaps 6;

Qy 18 RQMDDAAGRHHFASQCPRPTSPVSTDS---NMSAVVIQKAPAKKQKQPG----- 66
 Db 2294 rrrrsgggga---lrrapgpplaqdsqgrpsvclaraarpagqrlilpprtgqapr 2350
 Qy 67 -----HLRGRAYADLLPPFPVPPPAIKSPVQSKAQLVFPVMVPLASIBARTDR 117
 Db 2351 arlpqtkarsvqrerrgvlslppp-pppgelap-----rahpartprpgqdserr 2401
 Qy 118 SSDR-----KGG 124
 Db 2402 ggrvrtasagkyg 2414

Search completed: January 22, 2001, 12:19:50
 Job time: 1747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:28 ; Search time 325.28 Seconds
(without alignments)
30.894 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAQVAAAAYAGLAKVARRQM.....REALDGQVLDLRTNPSDPR 148

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Archived: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB ID	Description
1	761	100.0 1612 2	T30805
2	719	94.5 1651 2	T14160
3	101	13.3 1344 2	T14316
4	99	13.0 739 2	T21431
5	97.5	12.8 308 2	A38582
6	91	12.0 825 2	T27852
7	89	11.7 1882 1	GNVVT
8	88.5	11.6 821 2	S39983
9	87	11.4 826 2	A60385
10	86.5	11.4 891 2	T22560
11	85.5	11.2 662 2	T23757
12	85.5	11.2 1870 2	S37671
13	85.5	11.2 1872 2	S36152
14	85	11.2 300 2	T49225
15	85	11.2 698 2	T32594
16	85	11.2 716 2	T26998
17	85	11.2 1541 2	T02831
18	85	11.2 2424 2	T46480
19	85	11.2 2424 2	T46479
20	83.5	11.0 282 2	S53502
21	83.5	11.0 365 2	T24955
22	83	10.9 1336 2	S25716
23	83	10.9 1560 2	T47277
24	82.5	10.8 1013 2	T46422
25	82.5	10.8 1102 2	JC6316
26	82.5	10.8 1547 2	T28657
27	82	10.8 363 2	T16755
28	82	10.8 601 2	S56144
29	82	10.8 797 1	T46044

30	82	10.8	867 2	JH0225	L96 protein - Tipu
31	81.5	10.7	537 2	A46611	myosin-binding pro
32	81.5	10.7	555 2	I53869	zinc finger protei
33	81.5	10.7	555 2	T30349	structural protein
34	81.5	10.7	806 2	T35640	probable sensor ki
35	81.5	10.7	822 2	I38728	epidermal growth f
36	81.5	10.7	963 2	T04002	hypothetical prote
37	81.5	10.7	975 2	T08606	protein phosphatas
38	81.5	10.7	1603 2	S17983	gene posterior sex
39	80.5	10.6	197 2	F82029	probable periplasm
40	80.5	10.6	417 2	A72078	ct005 hypothetical
41	80.5	10.6	417 2	B81590	conserved hypothet
42	80.5	10.6	548 2	S59133	ETS2 repressor fac
43	80.5	10.6	2142 2	B35098	MBC class III hist
44	80	10.5	202 2	T11744	dehydrin - kidney
45	80	10.5	383 1	A48222	dematin 48K chain

ALIGNMENTS

RESULT 1
T30805
dutt1 protein - mouse
N:Alternate names: transmembrane receptor protein Robol homolog
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30805
R:Wu, M.C.; Lowe, N.; Fordham, R.; Rabbits, P.
submitted to the EMBL Data Library, July 1998
A:Description: The mouse homologue of human DUT11/H-robol gene: protein sequence and
A:Reference number: T20879
A:Accession: T30805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1612 <NID>
A:Cross-references: EMBL:T17793; NID:el329712; PID:el329713; PIDM:CAA76850.1
A:Experimental source: brain
C:Genetics:
A:Gene: dutt1
A:Map position: 16

Query Match 100.0%; Score 761; DB 2; Length 1612;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAQVAAAAYAGLAKVARRQM...QADAGRRHFHASCPRPTSPVSTDSNMSAVIQARPAKK 60

Db 1365 AQAQVAAAAYAGLAKVARRQM...QADAGRRHFHASCPRPTSPVSTDSNMSAVIQARPAKK 1424

Qy 61 QKQGHGLRRAYADLPPPPVPPPAIKSPVQSKAQLEVRPMVKPLASIEARTDRSSD 120

Db 1425 QKQGHGLRRAYADLPPPPVPPPAIKSPVQSKAQLEVRPMVKPLASIEARTDRSSD 1484

Qy 121 RKGSGYKRGREALDGRQVLDLRTNPSDPR 148

Db 1485 RKGSGYKRGREALDGRQVLDLRTNPSDPR 1512

RESULT 2
T14160
transmembrane receptor protein Robol - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14160
R:Kidd, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodenough
Cell 92, 205-215, 1998
A>Title: Roundabout controls axon crossing of the CNS midline and defines a novel class
A:Reference number: T17897; MUID:98117249
A:Accession: T14160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

RESULT 6
T27852
hypothetical protein ZK418.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 15-Oct-1999

0/255: 25 000 1000 frequency_0123456789 25 000 1000 1000000000 25 000 1000


```

Qy  143 NPSDPR 148
      :| | :
Db  234 SPEDKK 239

```

```

RESULT 12
S37671
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
C:Accession: S37671
R:Boigueleret, L.
submitted to the EMBL Data Library, August 1992
A:Reference number: S37671
A:Accession: S37671
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1870 <R0D>
A:Cross-references: EMBL: L215025; NID: g29374; PID: g29375
C:Genetics:
A:Map position: 6p21.3
A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;
C:Superfamily: collagen alpha 1(IV) chain

```

Query Match 11.2%; Score 85.5; DB 2; Length 1870;
Best Local Similarity 25.7%; Pred. No. 19;
Matches 39; Conservative 14; Mismatches 58; Indels 41; Gaps 7;

QY	18	QWMDARRRHHFASQCPPTSPVSTDS	---NMSAVIQKARPAKQKHQPGHREAYD	76	
DB	656	QQQQQHQQQQGQSSAPTPVPVSPQPV	LTGLNPAKAPKPPPKALYTGALGR	709	
QY	77	LPVVPPV	---PAIKSPVTSQKALQVLRPV	---MVKLASIARTDRSS	110
DB	710	-----PPPPMNFDFRNMIMPTVPRLL	QGRPPDLPTVPGVHSGVLPV	-----ERS	758
QY	120	DKKGGSTKRGKDALQWQLDR	---TMPSPD	148	
DB	759	DSRGSL	---SEPPDRHAPALREGTVPVQK	787	

```

RESULT 13
S36152
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu
C:Species: Homo sapiens (man)
C:Date: 06-Jun-1995 [sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
C:Accession: S36152
R:Irish, F.J.M.; Bouguerel, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Ju
Nature Genet. 3: 137-145, 1993
A:Title: Dense AU clustering and a potential new member of the NFkappaB family with
A:Reference number: S36152; MIMD:93272029
A:Accession: S36152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1872 <IR>
A:Cross-references: EMBL:125025
A:Note: in the authors' translation residues 32-34 are shown after residue 4 and, con
A:Note: the authors translated the codon AAT for residue 1000 as His
C:Genetics:
A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;
C:Superfamily: collagen alpha 1(IV) chain

```

Query Match 11.2%; Score 85.5; DB 2; Length 1872;
Best Local Similarity 25.7%; Pred. No. 19;
Matches 39; Conservative 14; Mismatches 58; Indels 41; Gaps 7;

Oy 18 RQMQDAAGRRHFHASQCPRPTSPVSTDS-NMSAVVIKARPAKKQKHQPGHLRREAYADD 76
:
Db 657 KOOOHHOOHCOGSAPPTVPVPPSPPPOVTLGAVPAPKAPPPPKALYPGALGR----- 710

[illegible]

Search completed: January 22, 2001, 12:27:33
Job time: 2130 sec

Query Match 11.2%; Score 85; DB 2; Length 698;
Best Local Similarity 31.0%; Pred. No. 7.3;
Matches 35; Conservative 10; Mismatches 48; Indels 20; Gaps 5;


```

Query Match      12.8%; Score 97.5; DB 1; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.25;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;

QY      3 AVAAAAETAGLKVRRQMDAGRHRHFSQCPRPTSPVSTDSNNSAVVIQKAPKAKQ 62
      || | | | | | | | | | | | | | | | | | | | | | | | |
Db      39 ATPAATPAGGREGEDDRBAAAGAGROBLASROPPWPLELTP----- 78

```

Qy 133 DGRQVTDLRINPSPD 147
| | || |
Db 311 PCGEGLTLRAKPPP 325

us-09-540-245a-20.rsp

HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC
CC *- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to licences@isb-sib.ch).
CC -----
DR EMBL: X57477; CAA040715.1; -.
DR EMBL: X57689; CAA040872.1; -.
DR EMBL: X57476; CAA040714.1; -.
DR EMBL: X57688; CAA040871.1; -.
DR INTERPRO: IPR000636; -.
DR INTERPRO: IPR020077; -.
DR PFM: PFM020; Ion_trans: 4.
DR PATINTS: FRO0157; CACCHANNEL.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Alternative splicing.
FT REPEAT 85 363 1.
FT REPEAT 473- 717 11.
FT REPEAT 1240 1523 111.
FT REPEAT 1560 1823 11V.
FT DOMAIN 1 98 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 99 117 S1 OF REPEAT I (POTENTIAL).
FT DOMAIN 118- 135 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 136 155 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 156- 167 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 168 185 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 186 190 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 191- 209 S4 OF REPEAT I (POTENTIAL).
FT DOMAIN 210 228 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 229 248 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 249 335 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 336- 360 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 361 487 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 488 506 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 507 521 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 522 541 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 542 549 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 550 568 S3 OF REPEAT II (POTENTIAL).
FT DOMAIN 569 578 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 579 597 S4 OF REPEAT II (POTENTIAL).
FT DOMAIN 598 616 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 617 636 S5 OF REPEAT II (POTENTIAL).
FT DOMAIN 637 689 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 690 714 S6 OF REPEAT II (POTENTIAL).
FT DOMAIN 715 1253 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1254 1272 S1 OF REPEAT III (POTENTIAL).
FT DOMAIN 1273 1288 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1289 1308 S2 OF REPEAT III (POTENTIAL).
FT DOMAIN 1309 1320 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1321 1339 S3 OF REPEAT III (POTENTIAL).
FT DOMAIN 1340 1350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1351 1369 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 1370 1388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1389 1408 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 1409 1495 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1496 1520 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1521 1575 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1576 1604 S1 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1605 1609 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1610 1629 S2 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1630 1637 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1638 1656 S3 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1657 1665 EXTRACELLULAR (POTENTIAL).

```

RESULT 7
IE68_HSV2
ID IE68_HSV2 STANDARD; PRT; 197 AA.
AC P14379;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)

```

```

RESULTS      8
SOS1_MOUSE
ID   SOS1_MOUSE  STANDARD;          PRT;   1319 AA.
AC   Q62245; Q62244;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SON OF SEVENTEENLESS PROTEIN HOMOLOG 1 (SOS-1) (MSOS-1).
GN   SOS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS; TISSUE=EYE;
RX   MEDLINE=92335328; PubMed=1631150;
RA   Bowtell D., Pu P., Simon M., Senior P.;
RT   "Identification of murine homologues of the Drosophila son of
RT   sevenless gene: potential activators of ras.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
RN   [2]
RP   STRUCTURE BY XRD OF 415-548.
RX   MEDLINE=97360234; PubMed=9217262;
RA   Koshiba S., Kigawa T., Kim J.-H., Shirouzu M., Bowtell D.,
RA   Yokoyama S.;
RT   "The solution structure of the pleckstrin homology domain of mouse
RT   Son-of-sevenless 1 (mSos1).";
RL   J. Mol. Biol. 269:579-591(1997).
CC   -/- PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC   SIMILIARITY).

```

```

RESULT 10
FURL_BOVIN
ID FURL_BOVIN STANDARD: PRI: 797 AA.
AC Q28183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 38, Last annotation update)
DE FURIN PRECURSOR (EC 3.4.21.75) (PAIRED BASIC AMINO ACID RESIDUE
DE CLEAVING ENZYME) (PACE) (DIBASIC PROCESSING ENZYME) (TRANS GOLGI
DE NETWORK PROTEASE FURIN).
GN PACE OR FUR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY:

```

```

SITE          752      765      CELL SURFACE SIGNAL.
FT SITE       752      765      TRANS GOLGI NETWORK SIGNAL.
SQ SEQUENCE   757 AA;  87250 MW;  466F28EC0446C3D2 CRC64;

Query Match          10.88; Score 82; DB 1; Length 797;
Best Local Similarity 27.14; Pred. No. 12;
Matches 23; Conservative 17; Mismatches 33; Indels 12; Gaps 3;

QY 30 HAS----QCPRPTSPVSTDSNMSAVIQKAPAKKHQHPGHLREAYADLPPPPVPPV 95
   ||| ||| ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 645 HASCATCGGAPATDCLSCPSHASLDPVVQTCRSQSQS-----SRESHQQQPPPRPPT 698

QY 86 AIKSPVTSQKQLEVRVWPKLAS 110
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 699 A--ZYATEPRLRADLLPSHLPEVVA 721

RESULT 11
VL96_IVR1
ID VL96_IVR1 STANDARD; PRT; 867 AA.
AC P22856;
DT 01-AGO-1991 (Rel. 19, Created)
DT 01-AGO-1991 (Rel. 19, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE L36 PROTEIN.
GN L36.
OS Tipula iridescent virus (TIV) (Insect iridescent virus 1).
OS Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91078645; PubMed=1701750;
RA Home W.A., Tajbakhsh S., Selly V.L.;
RT "Molecular cloning and characterization of a late Tipula iridescent
RT virus gene."
RL Gene 94:243-248(1990).
CC -/- FUNCTION: MAY BE INVOLVED IN TIV GENOMIC DNA PACKAGING IN A
CC MANNER RELATED TO THE GAG POLYPROTEINS OF THE MAMMALIAN VIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENSL; M2593; AAA7919.1; -.
DR PIR; JH0225; JH0225.
KW Repeat; DNA packaging; DNA-binding.
FT DOMAIN 697 867 RICH IN HYDROPHOBIC RESIDUES.
SQ SEQUENCE 867 AA; 96011 MW; F19DBDBE8F5CA103 CRC64;

Query Match          10.88; Score 82; DB 1; Length 867;
Best Local Similarity 25.48; Pred. No. 13;
Matches 34; Conservative 21; Mismatches 57; Indels 22; Gaps 6;

QY 35 PRPTSP-----VSTDNSMSAVIQK--ARPAKKHQHPGHLREAYADLPPPPVPPV 85
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 395 PRSSSPKPTKDLVITDSDEGIVKKITRKPSSPRSPPASVYR--SRTSPVSKSPASRP 453

QY 86 AIKSPVTSQK-----AQLEVR---PVMVPKLASIARTDRSSDR-KGSGYKRGREALD 133
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 454 RSKSPSVRAEITLDDGETPPSPVRPKSPVRPKSPSVVRPKSPSPVRPKSPSPASRPSPSKSP 513

QY 134 GRQVTDLRNTSPDP 147
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 514 SVRSKSPSPVRPKSP 527

RESULT 12
MYPH.CHICK

```


RESULT 15
ERF HUMAN

AC REF_HUMAN STANDARD; PRT; 548 AA.
 AC P50548;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ETS-DOMAIN TRANSCRIPTION FACTOR ERF.
 GN ERF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:96030784; PubMed-7586608;
 RA SOURCES D.N. Athanasios N.A., Beal G.J. Jr., Fisher R.J., Blair D.G.,
 RA Mavrothalassitis G.J.;
 RT "ERF: an ETS domain protein with strong transcriptional repressor
 RT activity, can suppress ets-associated tumorigenesis and is regulated
 RT by phosphorylation during cell cycle and mitogenic stimulation.";
 RL EMBO J. 14:4781-4793(1995).
 CC -1- FUNCTION: POTENT TRANSCRIPTIONAL REPRESSOR THAT BINDS TO THE H1
 CC ELEMENT OF THE ETS2 PROMOTER. MAY REGULATE OTHER GENES INVOLVED
 CC IN CELLULAR PROLIFERATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEOLAR.
 CC -1- PTM: PHOSPHORYLATED BY MULTIPLE KINASES INCLUDING PROBABLY ERK2.
 CC PHOSPHORYLATION REGULATES THE ACTIVITY OF ERF.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U15655; AAA86686.1; -.
 DR HSP: Q01543. 1FL1.
 DR INTERPRO: IPR000418; -.
 DR PFAM: PF00178; Ets; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Phosphorylation.
 FT DNA_BIND 27 107 ETS-DOMAIN.
 FT DOMAIN 166 171 POLY-SER.
 FT DOMAIN 290 293 POLY-GLY.
 FT DOMAIN 362 373 POLY-SER.
 FT DOMAIN 418 423 POLY-PRO.
 FT DOMAIN 496 499 POLY-GLY.
 FT MCD_RES 526 526 PHOSPHORYLATION (BY ERK2).
 FT MITAGEN 526 526 T-A; LOSS OF A PHOSPHORYLATION SITE.
 SQ SEQUENCE 548 AA: 58776 MW: C32A155394B1EEDD C8C64;

Query Match 10.6%; Score 80.5; DB 1; Length 548;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 39; Conservative 18; Mismatches 54; Indels 57; Gaps 7

QY	6	AAAYTAGLVKVAROMDAGRRHFHSAQSPRC-----TSVP-----STDSNMAVIOK	54
Db	330	AFLLHPLGLVLPQQRPD-----KCLPLPMAPMETPPVPSASSSSSSSSSPFFKFL	379
QY	55	ARPAKQKQKQGHILREAYAD-----LPPPPVPPAIPKSTVQSQAQLEV	100
Db	380	QRPLLEKQGRANGKAVAADKSGSAGLACGAGALAPPPPP-----QIKV	427
QY	101	PPVPPKIASLEARTDSRSGKSYKRGALRGDPLTRPSPDR	148
Db	428	EPISGEGSEYEV-TDIDSE-----PDGEGVEVTPDAPAPPK	465

Search completed: January 22, 2001, 12:29:56
Job time: 1297 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:54:12 ; Search time 559.88 Seconds
(without alignments)
30.983 Million cell updates/sec

Title: US-09-540-245a-20
Perfect score: 761
Sequence: 1 AQAVALAAAEYAGLVARRQM.....REALDGRVDTLRNPSDPDR 148

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Aligned: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTRMBL15.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	761	100.0	1612	11	089026	089026 mus musculus
2	719	94.5	1651	11	055005	055005 rattus norv
3	691	90.8	1651	4	Q9Y6N7	Q9Y6N7 homo sapien
4	101	13.3	1344	11	Q9Z214	Q9Z214 mus musculus
5	99	13.0	739	5	045408	045408 caenorhabdi
6	97.5	12.8	413	12	P89474	P89474 herpes simp
7	97.5	12.8	1262	5	Q9VWC2	Q9VWC2 drosophila
8	91	12.0	825	5	Q23484	Q23484 caenorhabdi
9	91	12.0	1017	10	Q9SS68	Q9SS68 arabidopsis
10	91	12.0	3080	5	Q9VRY3	Q9VRY3 drosophila
11	90.5	11.9	410	13	Q9IA21	Q9IA21 heterodontu
12	90	11.8	1212	10	Q9LGT8	Q9LGT8 oryza sativ
13	89	11.7	569	4	Q9UG03	Q9UG03 homo sapien
14	89	11.7	975	5	Q9VLA7	Q9VLA7 drosophila
15	89	11.7	2936	5	Q9NKP7	Q9NKP7 leishmania
16	88	11.6	811	5	Q9V8T0	Q9V8T0 drosophila
17	88	11.6	831	5	Q9NFI3	Q9NFI3 drosophila
18	88	11.6	980	11	P97837	P97837 rattus norv
19	87.5	11.5	621	10	Q9LW19	Q9LW19 arabidopsis

20	87	11.4	764	5	Q24708	Q24708 drosophila
21	86.5	11.4	296	10	Q40225	Q40225 lillium long
22	86.5	11.4	891	5	Q93763	Q93763 caenorhabdi
23	86	11.3	1650	5	Q9NBS4	Q9NBS4 caenorhabdi
24	85.5	11.2	604	10	Q9LW37	Q9LW37 arabidopsis
25	85.5	11.2	662	5	Q21536	Q21536 caenorhabdi
26	85	11.2	300	10	Q9LX35	Q9LX35 arabidopsis
27	85	11.2	698	5	Q44447	Q44447 caenorhabdi
28	85	11.2	716	5	Q9U2A6	Q9U2A6 caenorhabdi
29	85	11.2	1541	5	Q15837	Q15837 leishmania
30	84.5	11.1	2089	4	Q14676	Q14676 homo sapien
31	84.5	11.1	2089	4	Q9Y2W9	Q9Y2W9 homo sapien
32	84.5	11.1	2089	4	Q9UNU8	Q9UNU8 homo sapien
33	84.5	11.1	3201	5	Q9W0U2	Q9W0U2 drosophila
34	84.5	11.1	4880	11	Q9JLT1	Q9JLT1 rattus norv
35	84.5	11.1	5085	11	Q9JKS6	Q9JKS6 rattus norv
36	84	11.0	193	5	Q76215	Q76215 neospora ca
37	84	11.0	412	12	Q10415	Q10415 helicoverpa
38	84	11.0	754	4	Q9NUP6	Q9NUP6 homo sapien
39	83.5	11.0	262	10	Q40509	Q40509 nicotiana t
40	83.5	11.0	365	5	Q9XUP5	Q9XUP5 caenorhabdi
41	83.5	11.0	1180	5	Q9VRM2	Q9VRM2 drosophila
42	83.5	11.0	3232	5	Q9VFE8	Q9VFE8 drosophila
43	83	10.9	495	4	Q9NV63	Q9NV63 homo sapien
44	83	10.9	520	4	Q9P2K3	Q9P2K3 homo sapien
45	83	10.9	787	3	Q94096	Q94096 pneumocysti

ALIGNMENTS

RESULT 1
089026
ID 089026 PRELIMINARY; PRT; 1612 AA.
AC 089026;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DUTTL PROTEIN.
GN ROBO1 OR DUTTL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10093;
RW [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Wu M.C., Lowe N., Fordham R., Rabbitts P.;
RT "The mouse homologue of human DUTTL/H-robo1 gene: protein sequence and
RT chromosomal location."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR ENGL: Y17793; CAA76850.1; .
DR HSP: P56276; ITLK.
DR MGD: MGI:1274781; Robol.
DR INTERPRO: IPR001777; .
DR INTERPRO: IPR003006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; Ig; 5.
SQ SEQUENCE 1612 AA; 176406 MW; 5P2988C54479684B CRC64;

Query Match 100.0%; Score 761; DB 11; Length 1612;
Best Local Similarity 100.0%; Pred. No. 2.5e-65;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAVALAAAEYAGLVARRQM.....REALDGRVDTLRNPSDPDR 148
Db 1355 AQAVALAAAEYAGLVARRQM.....REALDGRVDTLRNPSDPDR 1424
Qy 61 QKQPHGLRRRYADLLPPPPPPPAIKSPVQSKAQLEVPVMPKLIASRTDRSSD 120
Db 1425 QKQPHGLRRRYADLLPPPPPPPAIKSPVQSKAQLEVPVMPKLIASRTDRSSD 1484

Qy 121 RKGSGYKRGALDGRQVDTLRLNPSDPR 148
|||||
Db 1485 RKGSGYKRGALDGRQVDTLRLNPSDPR 1512

RESULT 2
Q55005
ID Q55005 PRELIMINARY; PRT; 1651 AA.
AC Q55005;
DT 01-JUN-1996 (TrEMBLrel. 06, Created)
DT 01-JUN-1996 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TRANSMEMBRANE RECEPTOR ROBO1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=SPINAL CORD;
CY MEDLINE=98117249; PubMed=9458045;
Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
Goodman C.S., Fear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL: AF041082; AAC39960.1; .
DR HSSP: P56276; 1TLK.
DR INTERPRO: IPRO01777; .
DR INTERPRO: IPRO03006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; ig; 5.
KW Transmembrane.
SQ SEQUENCE 1651 AA; 180746 MW; 8A2452DD46E186B7 CRC64;

Query Match 94.5%; Score 719; DB 11; Length 1651;
Best Local Similarity 93.24; Pred. No. 3e-61;
Matches 136; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Qy 1 AQAIAAAAYAGLVARRQMDAAGRRHFASQCPRTSPVSTDSNNSAAVIQKARPAKK 60
|||||
Db 1404 AQAIAAAAYAGLVARRQMDAAGRRHFASQCPRTSPVSTDSNNSAAVIQKARPAKK 1463
Qy 61 QKQPGHLLREAYTADLPPPPVPAIKSPVQSKAQLEVRVPMVKLASEARTDRSSD 120
|||||
Db 1464 QKQPGHLLREAYTADLPPPPVPAIKSPVQSKAQLEVRVPMVKLASEARTDRSSD 1523
Qy 121 RKGSGYKRGALDGRQVDTLRLNPSDPR 148
|||||
1524 RKGSGYKRGALDGRQVDTLRLNPSDPR 1551

RESULT 3
Q9Y6W7
ID Q9Y6W7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6W7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
Goodman C.S., Fear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";

RL Cell 92:205-215(1998).
DR EMBL: AF040990; AAC39575.1; .
DR HSSP: P56276; 1TLK.
DR INTERPRO: IPRO01777; .
DR INTERPRO: IPRO03006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; ig; 5.
SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 90.84; Score 691; DB 4; Length 1651;
Best Local Similarity 88.5%; Pred. No. 1.5e-58;
Matches 131; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 1 AQAIAAAAYAGLVARRQMDAAGRRHFASQCPRTSPVSTDSNNSAAVIQKARPAKK 60
|||||
Db 1404 AQAIAAAAYAGLVARRQMDAAGRRHFASQCPRTSPVSTDSNNSAAVIQKARPAKK 1463
Qy 61 QKQPGHLLREAYTADLPPPPVPAIKSPVQSKAQLEVRVPMVKLASEARTDRSSD 120
|||||
Db 1464 QKQPGHLLREAYTADLPPPPVPAIKSPVQSKAQLEVRVPMVKLASEARTDRSSD 1523
Qy 121 RKGSGYKRGALDGRQVDTLRLNPSDPR 148
|||||
Db 1524 RKGSGYKRGALDGRQVDTLRLNPSDPR 1551

RESULT 4
Q9Z214
ID Q9Z214 PRELIMINARY; PRT; 1344 AA.
AC Q9Z214;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE RIG-1 PROTEIN.
GN RBG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF060570; AAD11628.1; .
DR HSSP: P56276; 1TLK.
DR MGD: MGI:1343102; Rbg1.
DR INTERPRO: IPRO01777; .
DR INTERPRO: IPRO03006; .
DR PFAM: PF00041; fn3; 3.
DR PFAM: PF00047; ig; 5.
SQ SEQUENCE 1344 AA; 143439 MW; 8B006034AC49CFEA CRC64;

Query Match 13.3%; Score 101; DB 11; Length 1344;
Best Local Similarity 27.7%; Pred. No. 0.14;
Matches 38; Conservative 18; Mismatches 59; Indels 22; Gaps 5;
Qy 29 FHASQCPRTSPVSTDSNNSAAVIQ-----KARPAKKQKHGLHREAYTADLPPPP 80
|||||
Db 1191 FHASQCPRTSPVSTDSNNSAAVIQ-----KARPAKKQKHGLHREAYTADLPPPP 1248
Qy 61 PVPPPAIKSPVQSKAQLEVRVPMVKL-----ASIEARTDRSSDRKSGSYKRGAE 131
|||||
Db 1249 PLPPPELRLKALGSA--GSRQVHPPRARAQWEEGASGASRGPTSSQR-GHPDGKES 1305
Qy 132 LDGRQVDTLRLNPSDPR 148
: : : : :
Db 1306 QGSRGRLKACSPNSPQ 1322

RESULT 5
Q45408

```

RESULT      6
P89474
ID P89474      PRELIMINARY;      PRT;      413 AA.
AC P89474;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HERPES SIMPLEX VIRUS TYPE 2 (STRAIN HG52), COMPLETE GENOME.
GS US1.
OS Herpes simplex virus (type 2).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OC NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary

```

```
Query Match      12.8%; Score 97.5; DB 12; Length 413;  
Best Local Similarity 27.6%; Pred. No. 0.091;  
Matches : Conservative 12; Mismatches 32; Indels 27; Gaps 3;  
  
Qy   76 DLPP-----PVPVPPAIIK-----PTVQSQAQLVPRVPVKIASIERTDSS 119  
    |:||| |||||::|::|::|::|::|:  
Db    3 DIPPOPLNLTWNAPNPSPGSGRRRPNPSSSESCGFOTESSESSESSEDGAG 62  
  
Qy   120 DRKGSTYGGREALDGROWTDLR-----TPMSD 146  
    ||::|::|::|::|::|:  
Db    6 DLRGGRRESRELGGTYFLDIASSTTGTESSTGTSPSD 100
```

```

Q95WU2      7
Q95WU2
ID   Q95WU2      PRELIMINARY;          PRT; 1262 AA.
AC   Q95WU2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   C015619 PROTEIN.
GN   C015619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
EX   MEDLINE=107196005; PubMed=10731132;
RA   Adams M.D., Celsiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Ananietis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.C., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chamoc M., Pfeiffer B.D.

```

```

RESULT          9
Q98568
ID   Q98568          PRELIMINARY;          PRT;   1017 AA.
AC   Q98568;
DT   01-MAY-2000 (TrEMBL; 13, Created)
DT   01-MAY-2000 (TrEMBL; 13, Last sequence update)
DT   01-JUN-2000 (TrEMBL; 14, Last annotation update)
DE   PUTATIVE PHOSPHORIBOSYLTRANSFERASE.
GN   T12J13.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Bacteria; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxId=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA   Ronning C.M., Koo H., Fujii C.T., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White D., Niernan W.C., Fraser C.W.;
RL   "Arabidopsis thaliana chromosome III RA T12J13 genomic sequence."
RT   Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AC009327; AAP03465.1; -.
DR   INTERPRO: IPR000008.
DR   PFAM: PF00168; C2; 3.
DR   PROSITE: PS00004; C2 DOMAIN 2; 3.

```

```

RESULT      8
Q23484
ID   Q23484          PRELIMINARY;      PRT;   825 AA.
AC   Q23484;
DT   01-NOV-1996    (TrEMBL|Ref. 01, Created)
DT   01-NOV-1996    (TrEMBL|Ref. 01, Last sequence update)
DT   01-JUN-2000    (TrEMBL|Ref. 14, Last annotation update)
DE   COSMID ZK418.
GN   ZK418.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC   Rhabditidae; Pelodirinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BRISTOL N2;

```

DD 155 ESRQNAHQRISSSSSS+ESSAGERSFFGPFASRRKARIATISSRQLEVED 165

Qy 1 AQAIAAAEYAGLAVARQMDAAGRHHFASQCPRPTS-----PVSTDSNMSAVVIQK 54
 | | | : | : : | | : | : | | | | : | : | |
 Db 529 AYATQAQRQQTLMQOQQOQRMSQQKHESCPVPPGMYSQLPWSPTDLSGWSQ- 587
 Qy 55 ARPAAKQ-----KHQPGHLRREAYADLPPPPVPPPAIKSPVQSKAQLVPRVMPV 106
 | : | | | : : | : | : | | : | | : : | :
 Db 588 AQPSHAQVISAINQMYHQOQHQOQQLNQOQSNPQM-NPAYQFOQQQPPGQAKINPILNA 646
 Qy 107 KLASIEARTDRSSDRKGS 125
 | : | : | |
 Db 647 NLKQNAQFNMQTMQPPGS 665

RESULT 15

Q9NKP7

ID Q9NKP7 PRELIMINARY; PRT: 2936 AA.

AC Q9NKP7:

01-OCT-2000 (TrEMBLrel. 15, Created)

01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE L712.2.

GN L712.2.

OS Leishmania major.

OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI_TaxID=5664;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FRIEDLIN;

RA Myler P.J.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC005930; AAF39727.1;

SQ SEQUENCE 2936 AA; 305212 MW; BE689E280242FB5B CRC64;

Query Match 11.74; Score 89; DB 5; Length 2936;
 Best Local Similarity 23.44; Pred. No. 4.5;
 Matches 36; Conservative 25; Mismatches 67; Indels 26; Gaps 5;

Qy 1 AQAIAAAEYAGLAVARQMDAAGRHHFASQCPRPTSPVSTDSNN-----S 48
 | | | : | : : | | : | : | | | | : | : | |
 Db 477 AATATGAERQRRASPKRSRDIAGARAGASSASP-PTTALSTSPPLASPPVLSPAERS 535
 Qy 49 AVVIQKARPAKQKHQPGHLR--REAYADLPP-----PPPVPPPAIKSPVQSKAQL 99
 | : : : : | | | | | : | : | | : : | : :
 Db 536 AAVLRQLKQVQRRLAHGHLRGTAAGANPLIEVKAHSPAAAFSNSTRGPEZSCSPCDIS 595
 100 VRPMVWPKLASIEARTD----RSSDRKGSYKGR 129
 | | | : : : : | : | : | :
 Db 596 VAASSVYSTATSSSSSSAAMKRERHGSARRNR 629

Search completed: January 22, 2001, 12:54:19

Job time: 2060 sec

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINE(S) OR MARK(S) ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.